Vector NTI Advance™ 10 User’s Manual

Published by:
Invitrogen Corporation
1600 Faraday Avenue
Carlsbad, CA 92008
USA

www.invitrogen.com

Copyright © 2004, 2005 Invitrogen Corporation. All rights reserved. This book contains proprietary information of Invitrogen Corporation. No part of this document, including design, cover design, and icons, may be reproduced or transmitted in any form, by any means (electronic, photocopying, recording, or otherwise) without prior written agreement from Invitrogen Corporation.

The software described in this document is furnished under a license agreement. Invitrogen Corporation and its licensors retain all ownership rights to the software programs offered by Invitrogen Corporation and related documentation. Use of the software and related documentation is governed by the license agreement accompanying the software and applicable copyright law.

Vector NTI Advance and Gateway are registered trademarks of Invitrogen Corporation in the United States and other countries. Logos of Invitrogen Corporation are also trademarks registered in the United States and may be registered in other countries. Other product and brand names are trademarks of their respective owners.

Generated in the United States.

Invitrogen Corporation reserves the right to make and have made changes, without notice, both to this publication and to the product it describes. Information concerning products not manufactured or distributed by Invitrogen Corporation is provided without warranty or representation of any kind, and neither Invitrogen Corporation nor its affiliates will be liable for any damages.

This version of the Vector NTI Advance™ 10 User’s Manual was published in September 2005.

Invitrogen’s free technical support for Vector NTI software is available exclusively through the web. For more information, check out the eSupport section of the Vector NTI user community at http://www.invitrogen.com/vectorNTIcommunity.

To obtain personalized technical support by telephone or email, you must have an annual support contract. Academic and government users may purchase an Academic Vector NTI Support Contract through the User Community. Commercial users may purchase an Advanced Support Contract by contacting Invitrogen at bioinfosales@invitrogen.com (for the Americas) or eurobioinfoorders@invitrogen.com (for European/Middle Eastern/African/Asian Pacific customers).

For paid support, use the following contacts:

United States
Phone: 240-379-4240 or 877-357-3114 (Toll-free, U.S.)
E-mail: bioinfosupport@invitrogen.com

Europe, Middle East, Africa, Asian Pacific
Phone: +44 (0) 141 814 6350
Email: eurobioinfosupport@invitrogen.com

The Invitrogen Technical Support team may request that you run the Vector NTI System Information utility and email them the resulting file. Vector NTI System Information is a utility application, which gathers data about the system on which Vector NTI is installed and consolidates that information into one file. The system information file is helpful in the diagnosis and resolution of system-related Vector NTI problems.
Table of Contents at a Glance

Chapter 1
Introduction to Vector NTI Advance ...............................................................................1

Chapter 2
Overview of Vector NTI Advance ....................................................................................5

Chapter 3
The Vector NTI User Interface ..........................................................................................13

Chapter 4
Tutorial: Molecule Display Windows ............................................................................25

Chapter 5
Tutorial: Molecule Editing .............................................................................................37

Chapter 6
Tutorial: Formatting Molecule Graphics ........................................................................45

Chapter 7
Tutorial: Database Explorer ..........................................................................................55

Chapter 8
Tutorial: PCR Analysis and Primer Design ..................................................................65

Chapter 9
Tutorial: Molecule Construction ...................................................................................77

Chapter 10
Tutorial: Molecule Design ...........................................................................................85

Chapter 11
Tutorial: Advanced Molecule Design ...........................................................................93

Chapter 12
Tutorial: Gel Display Windows; Restriction Digestion ..................................................99

Chapter 13
Tutorial: Internet Connectivity and Tools .....................................................................107
Chapter 14
Tutorial: AlignX ............................................................................................................ 115

Chapter 15
Tutorial: Bioannotator ................................................................................................. 125

Chapter 16
Tutorial: ContigExpress .............................................................................................. 131

Chapter 17
Database Explorer ....................................................................................................... 163

Chapter 18
Molecule Display Windows; Formatting Graphics ................................................... 203

Chapter 19
Editing Molecule Content ......................................................................................... 233

Chapter 20
Primer Design: PCR, Sequencing and Hybridization ............................................... 245

Chapter 21
Search and Analysis Tools ......................................................................................... 283

Chapter 22
Molecule Construction ................................................................................................ 305

Chapter 23
Gateway and TOPO Cloning ....................................................................................... 321

Chapter 24
Molecule Design .......................................................................................................... 349

Chapter 25
Gel Display Windows .................................................................................................. 365

Chapter 26
Report Generation ....................................................................................................... 375

Chapter 27
Shared Databases ....................................................................................................... 383

Chapter 28
AlignX ........................................................................................................................... 389

Chapter 29
BioAnnotator ............................................................................................................... 415

Chapter 30
ContigExpress: Project Explorer ............................................................................... 431

Chapter 31
ContigExpress: Fragment and Contig Windows ...................................................... 461
Table of Contents

Chapter 1
Introduction to Vector NTI Advance .............................................................. 1
  Welcome ...................................................................................................................1
  Getting Started with Vector NTI Advance ...........................................................1
  The Vector NTI User ...............................................................................................2
  The Vector Advance 9.0 User’s Manual Format ..................................................2
  Conventions Used in this Manual ...........................................................................2

Chapter 2
Overview of Vector NTI Advance ................................................................. 5
  Introduction ...............................................................................................................5
  Vector NTI Database Explorer ..............................................................................5
    Organizing your Biological Objects ......................................................................6
    Database Search .....................................................................................................6
    Import/Export Functions .......................................................................................6
    Archives - A Tool For Sharing Database Information ........................................6
  Features and Functions of Vector NTI ................................................................ 7
    Creating New Molecules .......................................................................................7
    Creating New Enzymes, Oligos, and Gel Markers ............................................7
    Oligos and Primers ...............................................................................................7
    Editing Database Objects ....................................................................................7
    Parent-Descendant Connections .........................................................................7
    Generating Restriction Maps ..............................................................................7
  Molecule Display ...................................................................................................8
    Text Pane .............................................................................................................8
    Graphics Pane .....................................................................................................8
    Sequence Pane .....................................................................................................8
    Creating Molecule Documents ..........................................................................8
  Gel Display Windows ...........................................................................................8
    Creating Gel Display Windows ..........................................................................8
    Creating Gel Samples and Gel Markers ...........................................................8
    Running the Gel ....................................................................................................9
  Align X ..................................................................................................................9
  BioAnnotator .........................................................................................................9
  ContigExpress .......................................................................................................9
GenomBench ............................................................................................................9
Remote Services Tools ...........................................................................................10
   BLAST Search/BLAST Viewer .............................................................................10
   PubMed/Entrez Search, Citation Viewer and Citation Table of Database Explorer .................................................................................10
   Remote Database Sources Wizard ........................................................................10
Miscellaneous Tools .............................................................................................10
   Analysis Monitor ......................................................................................................10
   GCG Converter ...........................................................................................................11
   Matrix Editor ................................................................................................................11
   AlignX Blocks ..............................................................................................................11
   3D-Mol ........................................................................................................................11
   Oligo Files Importer ....................................................................................................11
   Tools Manager .............................................................................................................11
   License Manager .......................................................................................................11
   Vector NTI System Information .................................................................................11

Chapter 3
The Vector NTI User Interface .............................................................................13

Introduction .............................................................................................................13
Launching Vector NTI ............................................................................................13
Online Help .............................................................................................................13
The Application Workspace ...................................................................................14
   Viewer Display Window Description ........................................................................14
   Viewer Panes ................................................................................................................15
   Viewer Window Operations .......................................................................................15
   Viewer Display Panes .................................................................................................16
Miscellaneous Interface Features ............................................................................18
   Selection Techniques ................................................................................................18
   Copying Display Window Text, Graphics or Sequences Image To The Clipboard ........19
   Printing from Display Windows ................................................................................21
   Exporting Molecule Data ..........................................................................................22
   Pane Manager ..........................................................................................................22

Chapter 4
Tutorial: Molecule Display Windows ......................................................................25

Introduction .............................................................................................................25
1. Launch Vector NTI .............................................................................................25
2. Inspect the Vector NTI Workspace and Database Explorer Windows ..............26
3. Create and Inspect a Molecule Display Window for pBR322 .........................26
4. Arrange the Display Window Conveniently ......................................................27
5. Select a Fragment or a Feature in the Graphics Pane ......................................28
   Modifying the Selection ..........................................................................................28
   Selecting Features ..................................................................................................29
   Selection Techniques Summary for the Graphics Pane ........................................29
6. Examine pBR322’s Nucleotide Sequence ........................................................29
7. Work with pBR322’s Text Description .............................................................31
8. Link pBR322’s Text Pane to the Graphics and Sequence Panes .......................32
9. Print pBR322’s Text Description, Graphical Map, and Sequence ..................33
10. Create a Display Window for 41BB_HUMAN ...........................................33
11. Create a Display Window for 1B14_HUMAN ............................................35
12. Close the Display Windows and Exit Vector NTI ....................................36

Chapter 5
Tutorial: Molecule Editing ...........................................................................37
  Introduction ..................................................................................................37
  1. Launch Vector NTI ..................................................................................37
  2. Open a Display Window for pBR322 .....................................................37
  3. Edit pBR322’s General Data .................................................................38
  4. Edit My pBR322’s Sequence .................................................................39
  5. Revert to Database Version of My pBR322 ...........................................40
  6. Insert a New Sequence Fragment .........................................................40
  7. Edit the TC(R) Signal ...........................................................................41
  8. Delete the P2_P Signal and Add a New Signal ......................................42
  9. Change the Starting Coordinate of My pBR322 ....................................42
 10. Close the Display Window and Exit Vector NTI ..................................43

Chapter 6
Tutorial: Formatting Molecule Graphics ..................................................45
  Introduction ..............................................................................................45
  1. Launch Vector NTI ................................................................................45
  2. Open pBR322 in a New Display Window .............................................45
  3. Arrange the Display Window Conveniently .......................................46
  4. Change the Arrangement Setup ..........................................................46
  5. Change the Graphics Setup of CDS Signals .......................................46
  6. Turn on the Picture Editing Mode .........................................................48
  7. Turn the TC(R) Arrow to Blue Cross-Hatching ..................................48
  8. Enlarge the TC(R) Arrow .....................................................................49
  9. Format TC(R)’s Label ...........................................................................50
 10. Add a Text Annotation .........................................................................50
 11. Experiment with Styles ......................................................................51
 12. Save the pBR322 Molecule Display to a Molecule Document file ......52
 13. Use the Saved Style Sheet to Format Molecule Display Windows ......52
 14. Close the Display Window and Exit Vector NTI ................................53

Chapter 7
Tutorial: Database Explorer ......................................................................55
  Introduction .............................................................................................55
  1. Launch Vector NTI ...............................................................................55
  2. Inspect the Database Explorer Window ............................................56
Chapter 8
Tutorial: PCR Analysis and Primer Design ................................................. 65

Introduction ............................................................................................................. 65
1. Launch Vector NTI and Open a Molecule Display Window ....................... 65
2. Select the Region of CoE1 ............................................................................. 65
3. Prepare for PCR Analysis ................................................................................. 66
4. Perform PCR Analysis ..................................................................................... 71
5. Inspect the PCR Analysis Folder ..................................................................... 71
6. Perform Oligo Analysis on a PCR Sense Primer ............................................ 71
7. Order Custom Primers .................................................................................... 72
8. Save a PCR Primer to the Database ............................................................... 73
9. Add Restriction Enzyme Sites to the PCR Primers ....................................... 73
10. Inspect the New PCR Analysis Results .......................................................... 75
11. Save the PCR Product to the Database and Open a Display Window .......... 75
12. Arrange the View of the Product .................................................................. 75
13. Close Display Windows and Exit Vector NTI .............................................. 76

Chapter 9
Tutorial: Molecule Construction ..................................................................... 77

Introduction ............................................................................................................. 77
1. Launch Vector NTI ............................................................................................ 78
2. Create Display Windows for pBR322 and pUC19 ......................................... 78
3. Arrange the Display Windows .......................................................................... 78
4. Define the First Fragment (from pUC19) .......................................................... 78
5. Define the Second Fragment (from pBR322) .................................................... 80
6. Inspect the Goal List ......................................................................................... 81
7. Enter General Information About the New Molecule ..................................... 81
8. Attempt to Construct the New Molecule ......................................................... 82
9. Fill in the AvaI Site of the pBR322 Fragment to Make it Blunt ....................... 83
10. Inspect Your New Molecule ........................................................................... 84
11. Close the Display Windows and Exit Vector NTI ........................................... 84
Chapter 10
Tutorial: Molecule Design ......................................................... 85
Introduction .................................................................................. 85
1. Launch Vector NTI and Create Display Windows for pBR322 and pUC19 ........ 86
2. Arrange Display Windows ............................................................... 86
3. Define the Recipient Fragment (from pUC19) ............................................ 86
4. Define the Donor Fragment (from pBR322) .................................................. 87
5. Inspect the Goal Molecule Definition List ................................................... 87
6. Enter General Information for Your New Molecule ........................................ 88
7. Prepare to Design the New Molecule .......................................................... 88
8. Set Your Preferences for Molecule Design .................................................... 89
9. Design the New Molecule ............................................................................. 89
10. Inspect the New Molecule ........................................................................... 90
11. Inspect the Design Plan .............................................................................. 90
12. Print Out the Design Plan .......................................................................... 91

Chapter 11
Tutorial: Advanced Molecule Design ................................................ 93
Introduction ...................................................................................... 93
1. Launch Vector NTI, Open and Arrange Display Windows ............................... 93
First Design: Complicated Recipient .............................................................. 94
2. Define the Recipient Fragment ...................................................................... 94
3. Define the Donor Fragment ........................................................................... 94
4. Inspect the Goal Molecule Definition List ..................................................... 95
5. Enter General Information for Your New Molecule ........................................ 95
6. Prepare to Design the New Molecule ............................................................. 95
7. Set the Design Preferences ............................................................................ 95
8. Design the New Molecule ............................................................................ 96
9. Inspect the New Molecule ............................................................................ 96
10. Inspect the Graphical Map and Text Description of Your New Molecule ......... 96
Second Design: Complex Donor Fragment ...................................................... 97
1. Define Recipient ....................................................................................... 97
2. Define Donor ............................................................................................. 97
3. Inspect the Goal List .................................................................................... 98
4. Design the New Molecule .......................................................................... 98
5. Inspect and Print the New Molecule ............................................................. 98

Chapter 12
Tutorial: Gel Display Windows;
Restriction Digestion ........................................................................ 99
Introduction .......................................................................................... 99
1. Launch Vector NTI .............................................................................................99
2. Create a New Gel and Open a Display Window .................................................99
3. Create a Sample and Add it to the Gel ..............................................................100
4. Add Gel Markers ..............................................................................................101
5. Arrange the Gel Display Window Conveniently ..............................................102
6. Inspect the Gel Display Window Text Pane .....................................................102
7. Change Display Color for a Fragment ..............................................................102
8. Run the Gel .......................................................................................................103
9. Gel Animation ...................................................................................................103
10. Optimize the Gel Display ...............................................................................103
11. Change to a Polyacrylamide Gel from the Electrophoresis Profile ...............104
12. Estimate Fragment Separation Time ...............................................................104
13. Save the Gel Display Window to a Gel Document File .................................105
14. Close the Gel Display Window ......................................................................105
15. Calculate Restriction Fragments .....................................................................105

Chapter 13
Tutorial: Internet Connectivity and Tools ....................................................107
Introduction ...........................................................................................................107
1. Launch Vector NTI and Open pBR322 in a New Display Window ..........107
2. Select the Whole Sequence of pBR322 and Use the BLAST Search Tool ......107
3. Display the Result of the Query in a Molecule Display Window .................109
4. Save (Send to) or Open the Result of the Query Using the Save Button ......110
5. Use Alignment and Analysis Tools .................................................................110
6. Saving Search Results as Comments .............................................................112
7. Close the Molecule Display Windows and Exit Vector NTI ......................113

Chapter 14
Tutorial: AlignX ............................................................................................115
Introduction ...........................................................................................................115
1. Launch AlignX from the Database Explorer .................................................115
2. Examine AlignX Display Window .................................................................116
3. Add Molecule to Alignment .........................................................................118
4. Inspect the Guide Tree ...................................................................................118
5. Review the Alignment Parameters and Matrix .............................................118
6. Examine the Alignment Pane ....................................................................119
7. Edit the Alignment .........................................................................................120
8. Generate Dot Matrix Plot ............................................................................120
9. Export Alignment Molecules .......................................................................122
10. Copy and Print an AlignX Pane .................................................................122
11. Save the Project and Close AlignX .............................................................122
Chapter 15
Tutorial: Bioannotator ................................................................. 125

Introduction ..................................................................................... 125
1. In Vector NTI, Open a Molecule Display Window for pSV2-dhfr; Create a New Protein by Translating a CDS ................................................................. 125
2. Launch BioAnnotator from a Vector NTI Molecule Display Window .......... 126
3. Review the BioAnnotator Display Window ........................................... 127
4. Examine and Modify Plot Properties .................................................. 127
5. Modify Plot Layouts .............................................................................. 128
6. Format the DHFR Sequence ............................................................... 128
7. Open a DNA Project in BioAnnotator .................................................. 129
8. Add New Analyses to BioAnnotator ................................................... 129
9. Copy, Paste, and Print BioAnnotator Panes ........................................ 130
10. Close the Display Window and Exit Vector NTI ............................... 130

Chapter 16
Tutorial: ContigExpress ................................................................. 131

Introduction ..................................................................................... 131
Session 1 Project Explorer ..................................................................... 131
1. Open a ContigExpress Window ......................................................... 131
2. Open a ContigExpress Project and Add a Fragment to the Project ........... 132
3. Examine the ContigExpress Project Explorer Window ........................ 133
4. Create Assemblies of Contigs ............................................................. 133
   Start Assembling .................................................................................. 133
   Assembly Setup .................................................................................... 134
   Examine Assemble Results .................................................................... 135
5. Review and Modify Viewing Options ................................................. 136
6. View Contents in the Tree and Content Panes ..................................... 139
7. Manage Fragment Lists and Project Items ........................................... 139
8. Save and Close a ContigExpress Project ........................................... 140
Session 2 Working in Fragment Viewer ................................................ 140
1. Open and Inspect the Fragment Window ........................................... 140
   Text Pane ......................................................................................... 141
   Sequence Pane ................................................................................... 142
   Chromatogram Pane .......................................................................... 143
2. Sequence Selection Techniques ......................................................... 143
3. Display ORFs and Translate the Nucleotide Sequence ......................... 144
   Display ORFs for the Sequence ........................................................... 144
   Translate the Nucleotide Sequence ...................................................... 145
4. Edit the Nucleotide Sequence in the Sequence Pane ........................... 146
5. Edit a Sequence in the Chromatogram Pane ....................................... 149
6. Close the Fragment Window ............................................................. 149
Chapter 17

Database Explorer .................................................................163

Introduction ...............................................................................163

Opening Vector NTI Database Explorer .................................164

Elements of Vector NTI Database Explorer .............................165

Database Tables .......................................................................165

  Citation Table .......................................................................165

Database Subsets ......................................................................167

Database Objects ......................................................................167

  Database Object System Fields ............................................168

  User-defined Fields ...........................................................170

Database Search .......................................................................171

  DNA/RNA and Protein Molecule Database Searches ............172

  Enzyme Database Search ...................................................176

  Oligo Database Search .......................................................176

  Gel Marker Database Search .............................................177

  Citation Database Search ..................................................178

  BLAST Results and Analysis Results Database Searches ......178

Database Management Operations ........................................178

  Backup, Restore and Cleanup .............................................178

  Selecting the Local Database .............................................179

Operations on Subsets ............................................................179

  Creating a New Subset .......................................................179

  Editing a Subset ..................................................................180

  Rename Subset ....................................................................181

  Dismiss Subset ....................................................................181

  Clear Subset ........................................................................181

  Delete Contents ....................................................................181

  Subset Summary ...............................................................181

  Subset Properties ..............................................................181

  Import Molecules into a Subset .........................................181

  Import Enzymes into a Subset ............................................182

  Export Subset ......................................................................184

  Copy and Save or Print a List of the Objects in a Subset ......184

Operations on Database Objects ............................................185

  DNA/RNA Molecule Operations .........................................186

  Protein Molecule Operations .............................................186

  Enzyme Operations ..........................................................186
Chapter 19

Editing Molecule Content ....................................................... 233

Introduction .................................................................................. 233

Edit Operations in a Molecule Display Window .................................. 233

Molecule Editing ............................................................................. 233

Edit Operations in Molecule Editor .................................................. 237

The General Information Tab .............................................................. 238
The DNA/RNA Molecule Tab .............................................................. 239
The Sequence and Maps Tab (DNA/RNA) ............................................. 239
Sequence and Features Tab (Protein) .................................................... 242
User Fields Tab .................................................................................. 243
Comments Tab .................................................................................. 243
Keywords Tab ................................................................................... 244
Changing the Starting Coordinate of a DNA Molecule ......................... 244
Making the Reverse Complement of a DNA Molecule ........................ 244

Chapter 20

Primer Design: PCR, Sequencing and Hybridization ..................... 245

Introduction .................................................................................... 245

PCR Primers ..................................................................................... 246

Saving and Loading PCR Settings ..................................................... 247
Find PCR Primers Dialog Box ............................................................. 247
Amplify Selection Dialog Box .............................................................. 255
Amplify Features Dialog Box ............................................................... 255
Long PCR Dialog Box ....................................................................... 258

Primer Tab ....................................................................................... 259

PCR Using Existing Oligos Dialog Box ............................................. 259
Alignment PCR Analysis ................................................................. 260
Multiplex PCR Analysis ................................................................. 263
The PCR Analysis Folder ................................................................. 271
Operations with PCR Products .......................................................... 272
Operations with PCR Primers ............................................................. 272
Saving, Viewing and Printing PCR Analysis Results ......................... 273

Sequencing Primers ........................................................................ 275

Sequencing Primers Dialog Box ....................................................... 275
Primer Similarity Dialog Box ............................................................. 276
Primer/Oligo Parameters Dialog Box ................................................. 277
Primer/Oligo Quality Specifics Dialog Box .......................................... 278
The Sequencing Primers Folder .......................................................... 279

Hybridization Probes ........................................................................ 279
Chapter 21
Search and Analysis Tools ................................................................. 283

Introduction ........................................................................................................... 283
Oligonucleotide Analyses ..................................................................................... 283
The Oligo List ................................................................................................................ ........... 283
Order Custom Primers .......................................................................................................... .... 284
Oligo Editor Dialog Box ....................................................................................................... .... 285
Thermodynamic Properties Analysis (Oligo Analysis) Dialog Box ...................... 286
Dimers & Hairpin Loops Dialog Box ....................................................................................... 288
Oligo Duplexes Dialog Box ...................................................................................................... 289
ORFs and Motifs Search ............................................................................................... 290
ORFs Search Results ........................................................................................................... .... 290
Ultimate ORF Blast ............................................................................................................ ...... 290
Motifs Search Results ......................................................................................................... ...... 291
Mutagenesis .......................................................................................................... 291
Translation ............................................................................................................ 292
Translating in the Sequence Pane ............................................................................................. 293
Translating into a New Protein Molecule ................................................................................. 293
Setting the Genetic Code for DNA/RNA Translations ......................................................... 294
Back Translation ................................................................................................... 297
Opening the Back Translation Dialog Box ............................................................................... 297
Codon Usage Tables ............................................................................................................. 298
Translation Tables and Codon Ambiguity ................................................................................ 299
Back Translation Copy Options ................................................................................................ 299
Restriction Analyses ............................................................................................. 300
Restriction Fragments/Restriction Sites ................................................................................... 300
RFLP Analysis ................................................................................................................. ......... 300
Finding Common Non-Cutter Restriction Enzymes ................................................................. 302
Restriction Report ............................................................................................................ ............. 304

Chapter 22
Molecule Construction ................................................................. 305

Introduction ........................................................................................................... 305
Molecule Types: Basic And Constructed ............................................................... 305
Methods of creating new DNA/RNA molecules .................................................... 306
Creating New DNA/RNA or Protein Molecules Using Splicing .................................. 307
Tools for Creating New Constructed and Designed DNA/RNA Molecules .... 308
Describing Component Fragments in the Fragment Wizard .................................... 309
The Goal Molecule Definition List (Goal List) ........................................................... 311
Constructing a New Molecule .............................................................................. 312
The Construct & Design Molecule Dialog Boxes ......................................................... 312
Editing the Termini of Fragment of Molecule-type Component Fragments .............. 317
The Control Fields and Commands .................................................................................. 318
Reconstructing a Molecule ..................................................................................... 319
Chapter 23
Gateway and TOPO Cloning .................................................................321
   Introduction to Gateway Cloning ..........................................................321
   Creating Entry Clones ........................................................................323
      Generate an Entry Clone by Recombining an Expression Clone by BP 323
      Generate an Entry Clone by PCR Amplification and BP Recombination 324
      Generate an Entry Clone by BP Recombination ..................................329
      Generate an Entry Clone by PCR Amplification and Ligation into a pENTR Vector 329
   Obtaining a Functional Entry Clone from Invitrogen ...............................331
   Performing an LR Recombination .........................................................331
   Creating a New Destination Vector ....................................................335
   Introduction to TOPO Cloning ..............................................................337
   Creating a New Destination Vector ....................................................335
   High Throughput Operations ................................................................344

Chapter 24
Molecule Design ..................................................................................349
   Introduction ..........................................................................................349
   Creating a New DNA Molecule Using Design ....................................349
   Describing the Goal Molecule ..............................................................350
      Defining the Recipient Using Fragment Wizard ..................................350
      Defining the Donors Using Fragment Wizard .....................................353
   The Goal Molecule Definition List (Goal List) .....................................355
   Starting the Design ................................................................................355
      The Component Fragments Field ........................................................356
   Defining Design Fragments in Fragment Editor .................................356
   Design Parameters ...............................................................................358
   Design Preferences ...............................................................................359
      Other Preferences ...............................................................................360
      Priorities ............................................................................................360
   Advanced Design Preferences ............................................................360
   The Design Process ...............................................................................361
   Returning to the Pre-Design State .......................................................362
   The Design Description Folder ............................................................362
      Cloning Step Subfolders ....................................................................362
   Redesigning a Molecule .......................................................................364

Chapter 25
Gel Display Windows ..........................................................................365
   Introduction ..........................................................................................365
   Opening a Gel Display Window ..........................................................365
      Opening a New Gel Display Window .................................................366
      Opening a Gel Document File ............................................................367
   Elements of a Gel Display Window ....................................................367
Chapter 29
BioAnnotator ................................................................. 415

Introduction ........................................................................ 415
Launching BioAnnotator ....................................................... 415

Features of the BioAnnotator Workspace ................................ 415

Elements of BioAnnotator Display Windows ......................... 416
Copying a Raw Sequence ..................................................... 422

BioAnnotator Projects .......................................................... 422

Creating a New Project .......................................................... 422
Opening an Existing Project .................................................... 422
Analysis List Setup ............................................................... 423
Analysis Properties ............................................................... 424
Analysis Descriptions ...............................................................426
DNA/RNA Analyses .................................................................426
Protein Analyses .................................................................426
ProtScale Algorithm ...............................................................427
Advanced Analyses: BLOCKS, ProSite, PFAM and Proteolytic Cleavage ......428
Adding Custom Protein Analyses .................................................428
Other BioAnnotator Features ..................................................429

Chapter 30
ContigExpress: Project Explorer .................................................431

Introduction ........................................................................431
Launching ContigExpress .........................................................431
Features of the ContigExpress Workspace ......................................432
ContigExpress Project Explorer .....................................................432
  Elements of Project Explorer .................................................432
  Tree Pane .........................................................................433
  List Pane .........................................................................435
General Project Explorer Operations ...........................................437
  Opening or Creating a Project ...............................................437
  Importing Fragments to a Project ..........................................438
  Finding Fragments ............................................................440
  Selection Techniques ..........................................................442
  Assembling Selected Items ..................................................442
  Loading and Saving Assembly Setup Settings .........................444
  Assembly Tab .....................................................................444
  Clipping Tab .......................................................................445
  Overlap Tab .......................................................................447
  Contig Tab .........................................................................448
  Description Tab ...................................................................449
  Lite Settings Tab ................................................................450
  Registering Assembly Setup Changes .....................................451
  Viewing Project Item Properties ............................................452
  Copying Selected Fragments or Contigs .................................459
  Printing ............................................................................459
  Exporting Items from the Project ..........................................460
  Saving and Closing a Project ................................................460
  Exit Project Explorer ..........................................................460

Chapter 31
ContigExpress: Fragment and Contig Windows .........................461

Introduction ........................................................................461
Features of the ContigExpress Workspace ......................................461
  Switching Between Panes ....................................................462
  Status Bar .........................................................................462
General User Actions ...............................................................463
  Finding a Feature in Other Panes ..........................................463
  Editing Fragment or Contig Properties .................................463
  Saving a Current Window ....................................................464
  Closing a Current Window ..................................................464
Closing All Windows ................................................................. 464
Selecting All Data ...................................................................... 464
Deleting Data ........................................................................... 464
External Tools Usage ................................................................. 464
Fragment Viewer ........................................................................ 465
  Launching a Fragment Viewer ................................................... 465
  Features of a Fragment Viewer .................................................. 465
  Miscellaneous Operations in the Sequence Pane .......................... 468
  Fragment Viewer Operations .................................................... 470
  Window Viewing Options ......................................................... 472
Contig Viewer ............................................................................. 473
  Launching a Contig Viewer ........................................................ 473
  Features of a Contig Viewer ....................................................... 473
Common Operations in the Graphics and Contig Alignment Panes .... 485
  Highlighting a Fragment ............................................................ 485
  Moving a Fragment .................................................................. 485
  Reverse Complement a Fragment or Contig ............................... 486
  Declare a Reference Sequence .................................................. 486
  Use of a Reference Sequence .................................................... 486
  Window Viewing Options ......................................................... 487
  Copy and Camera Features ....................................................... 490

Chapter 32
ContigExpress: Group Operations on Fragments ......................... 493
  Introduction ............................................................................. 493
  Fragment Trimming ................................................................. 493
  Fragment Ends and Vector Contamination Trimmer Dialog Boxes .... 493

Chapter 33
GenomBench .............................................................................. 505
  Introduction ............................................................................. 505
  Launching GenomBench .......................................................... 506
    Configuring Database Paths .................................................... 506
    Opening a Sequence from a DAS Server or a Local Database ....... 509
    Loading Features into the Feature Map Pane .............................. 510
  Features of the GenomBench Window ...................................... 513
    General Pane Operations ....................................................... 513
    Overview Pane ....................................................................... 515
    Feature Map Pane ................................................................... 515
    Information Pane ..................................................................... 521
  Undo and Redo Operations ...................................................... 529
  Configuring GenomBench Parameters ...................................... 530
    Setting Feature Type Filters ................................................... 530
    Editing Graphic Styles .......................................................... 531
    Editing Fragment Loading Configurations ................................. 534
  GenomBench Tools .................................................................. 539
    Feature Search Tool .............................................................. 539
    Pattern Search Tool .............................................................. 540
Sim4 and Spidey Alignment Tools .......................................................... 542
Querying Public Databases ....................................................................... 544
Sending Fragments to Vector NTI ............................................................ 551

Saving Sequences .................................................................................... 552
Saving Fragments ...................................................................................... 552
Creating New Sequences .......................................................................... 552

Importing and Exporting Sequences ........................................................ 554
Importing Files into GenomBench ............................................................. 554
Exporting Sequences to GenBank Files .................................................. 556

Printing in GenomBench ........................................................................ 556
Copying in GenomBench .......................................................................... 557
Copying by Camera ................................................................................... 557
Copying by the Copy Command ............................................................... 558

Chapter 34
Analysis Monitor - Advanced Analysis Tools ........................................ 559

Introduction ............................................................................................. 559
PFAM, ProSite and BLOCKS Databases .................................................. 559
Launching Advanced Analyses .................................................................. 560
Elements of the Analysis Monitor Workspace ........................................ 561
Configuring Advanced Analysis Parameters ......................................... 562
Sim4 ........................................................................................................... 562
Spidey ....................................................................................................... 564
BLOCKS ...................................................................................................... 565
PFAM .......................................................................................................... 565
Proteolytic Cleavage .................................................................................. 566
ProSite ....................................................................................................... 568
BLAST Search in Analysis Monitor ........................................................ 569

Operations on Analyses .......................................................................... 570
Starting an Analysis .................................................................................. 570
Stopping an Analysis ............................................................................... 570
Editing, Copying and Deleting ................................................................. 570
Saving ........................................................................................................ 571
Sending to Other Vector NTI Advance Applications ............................. 571

Viewing Analysis Results ........................................................................ 571
Viewing Analysis Results in Vector NTI ................................................ 571
Viewing Analysis Results in GenomBench .............................................. 572

Chapter 35
BLAST Search and BLAST Viewer ......................................................... 575

Introduction ............................................................................................. 575
BLAST Search Dialog Box ...................................................................... 575
Database Selection Parameters ................................................................. 576
Sequence/Parameters Pane Tabs ............................................................... 578
BLAST Options ......................................................................................... 582

BLAST Search Results .......................................................................... 582
BLAST Viewer .......................................................................................... 583
Chapter 36

AlignX Blocks ........................................................................................................... 593

Introduction ........................................................................................................ 593

Opening AlignX Blocks: ..................................................................................... 593

Features of the AlignX Blocks Workspace ...................................................... 593

Text Pane ........................................................................................................... 594
Block List Pane .................................................................................................. 595
Schematic Pane .................................................................................................. 597
Alignment Pane .................................................................................................. 597

AlignX Blocks Projects ...................................................................................... 598

Opening an Existing Project ............................................................................... 598
Creating a New Project ...................................................................................... 598
Adding Molecules to the Project ...................................................................... 598
Block Search ...................................................................................................... 598
Adding Sequences to Block Sequence Space ............................................... 599
Defining a Block ................................................................................................. 599
Linking and Unlinking ....................................................................................... 599
Color Setup ....................................................................................................... 599

Alignment Pane .................................................................................................. 600
Deleting a Molecule ............................................................................................ 600

AlignX Blocks Parameters ............................................................................... 601

Parameters Tab: ................................................................................................ 601
Blocks’ Color Tab: .............................................................................................. 601
Score Matrix Tab ................................................................................................. 602

Templates ............................................................................................................ 602

Other AlignX Blocks Features ........................................................................... 603

Chapter 37

3D-Mol .................................................................................................................. 605

Introduction ....................................................................................................... 605

3D-Mol Supported File Formats ...................................................................... 605

Downloading 3D Structure Files ...................................................................... 605

Opening 3D-Mol .................................................................................................. 606

Elements of 3D-Mol .......................................................................................... 606

Project Pane ...................................................................................................... 607
Structure Pane .................................................................................................. 609
Graphics Pane .................................................................................................. 610
Sequence Pane .................................................................................................. 611

Graphics Pane 3D Options Dialog Box ........................................................... 616
Stereo Viewing Option ....................................................................................... 620

Sequence Pane Viewing Options ...................................................................... 620

3D-Mol Tools ..................................................................................................... 620
Connolly Surface Calculation ................................................................. 661
Varshney Surface Calculation ................................................................. 661
AlignX .................................................................................................... 661
Basic Alignment Algorithms ................................................................. 661
Substitution Matrices ........................................................................... 662
Phylogenetic Tree .................................................................................. 662
Alignments and Similarity .................................................................... 662
BioAnnotator ....................................................................................... 662
DNA/RNA Analyses Citations ............................................................... 662
Protein Analyses (ProtScale Analyses) ................................................... 663
Properties Plots .................................................................................... 663
ContigExpress ..................................................................................... 665
Contig Assembly .................................................................................. 665
BLAST .................................................................................................. 665
Miscellaneous ..................................................................................... 665
Molar Extinction Coefficient ................................................................. 665
Oligo Analysis ..................................................................................... 665
PCR Analysis ...................................................................................... 666
General Bioinformatics ........................................................................ 666
Biostatistics ........................................................................................ 666
REBASE ............................................................................................. 666

Appendix C
ASCII Format; IUB Codes (IUPAC CODES) ......................................... 667
Format for ASCII Sequence Files .......................................................... 667
IUPAC Formats recognized by Vector NTI ............................................ 667

Appendix D
Parent-Descendant Relationships in Vector NTI .................................... 669
Overview of Parent-Descendant Relationships ...................................... 669
Viewing Parent-Descendant Relationships ........................................... 670
Managing Parent-Descendant Relationships ....................................... 671
  Modifying a Descendant ................................................................. 671
  Modifying a Parent ......................................................................... 672
  Deleting a Parent from the Database ............................................... 673
  Deleting a Descendant from the Database ....................................... 673

Appendix E
Tm Calculations in Vector NTI .............................................................. 675
General Information ........................................................................... 675
  Usefulness of Thermodynamic Tm Versus %GC Tm ......................... 675
  Effects of Primer (Probe) and Salt Concentration on Tm Calculations ... 676
%GC Tm Calculation .......................................................................... 676
Thermodynamic Tm Calculation ........................................................ 676
  Oligos Containing IUB Ambiguity Characters .................................. 679
  RNA Oligos .................................................................................... 679
Primer/Probe Tm, TaOpt and Similarity Calculations ............................ 680
COMMERCIAL USERS: IN ORDER TO ACCESS THE SOFTWARE YOU MUST ENTER INTO THE SOFTWARE PRODUCT LICENSE AGREEMENT IN SECTION I.

ACADEMIC OR NOT-FOR-PROFIT USERS: IN ORDER TO ACCESS THE SOFTWARE YOU MUST ENTER INTO THE ACADEMIC OR NOT-FOR-PROFIT USER SOFTWARE LICENSE AGREEMENT IN SECTION II.

I. Software Product License Agreement Terms for Commercial Users

This is a legal agreement ("Agreement") between you, the party purchasing a product license, and Invitrogen Corporation, having offices at 1600 Faraday Avenue, Carlsbad California 92008 USA ("Licensor"). This Agreement is part of a package that includes one or more software products and certain electronic and/or written materials. This Agreement covers your licensing of the Software (as the term is defined below). BY CLICKING YOUR ACCEPTANCE OF THIS AGREEMENT, OR BY INSTALLING OR USING THE SOFTWARE OR ANY OTHER COMPONENT OF THE PACKAGE, YOU ACKNOWLEDGE THAT YOU HAVE READ ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT, UNDERSTAND THEM, AND AGREE TO BE LEGALLY BOUND BY THEM. If you do not agree to the terms of this Agreement, you may not install or use the Software.

As used in this Agreement: "Authorized Users" means, collectively, the personnel authorized by you to use the Software, for whom you have purchased a License (as defined below) and paid the corresponding license fees. Unless otherwise expressly allowed by this Agreement, Authorized Users may include only your employees having a need to know; Authorized Users may not be entities or persons in the business of licensing or otherwise providing products or services competitive with the Software; "Designated Site" means your facilities or offices located at the postal address provided to Licensor for your billing and invoicing purposes, unless otherwise indicated in a license key provided to you; and "Software" means the software product(s) accompanying this Agreement, including the associated user manuals, user documentation and application program interfaces, and license key(s) provided, and any patches, updates, upgrades, improvements, enhancements, fixes and revised versions of any of the foregoing that may be provided to you from time to time, and any combination of the foregoing.

1. YOUR PAYMENT OBLIGATIONS. You agree to pay all amounts due or incurred by you, including any late payment fees, as are specified in this Agreement or in an invoice. If any authority imposes a duty, tax or similar amount (other than based on Licensor’s income), you agree to pay, or to promptly reimburse Licensor for, all such amounts. You are also responsible for paying all shipping charges and Licensor’s other reasonable out-of-pocket expenses incurred in performing this Agreement. Unless otherwise indicated, all invoices are payable thirty (30) days from the date of invoice. Overdue amounts are subject to a late payment charge, at the lower rate of (i) one and one-half percent (1.5%) per month, or (ii) the maximum legal rate. You agree to promptly pay or reimburse Licensor for all costs and expenses, including all reasonable attorneys’ fees, related to breach of your obligations under this Agreement or Licensor’s enforcement of this Agreement. All shipments by Licensor or its designee are FCA point of shipment.

2. GRANT OF SOFTWARE LICENSE. Subject to the terms and conditions of this Agreement, Licensor grants to you a non-exclusive, non-transferable license (the "License") to use the Software for your internal operations and internal data processing purposes within and up to the Authorized Scope for which you have paid
the applicable fees (if any) and registered the Software for use. The term “Authorized Scope” means the following, and any other capacity or use restrictions on the license granted to you:

- If you have purchased a “Static” license, then, for each such license purchased you may install and operate the Software on a designated computer located at the Designated Site. You may not network the Software for use at any other location. Use of the Software is limited to the computer designated for each Static license purchased.

- If you have purchased a “Network” license, then you may install and operate the Software concurrently on up to the number of workstations for which you have purchased a license. All workstations must be at the Designated Site. You may not network the Software for use at any other location; neither may you enable operation of concurrent “instances” of the Software beyond the scope of the Network license purchased. For example, if you purchased a Network license permitting concurrent use on five computers, you may enable the Software for concurrent use on up to five workstations at the Designated Site, regardless of whether such computers are networked.

- If you have purchased a “Dynamic” license, then you may allow up to the number of users for which you have purchased a license to operate the Software concurrently on any of the designated computers for which you have purchased a “Dynamic license seat”. The computers must all be located at the Designated Site. You may not network the Software for use at any other location, or for use on any computers other than those for which you have purchased a Dynamic license seat.

- If you have purchased a “Corporate” license, then you may install and operate the Software concurrently on all workstations at the Designated Site(s) for which you have purchased such license. You may not network the Software for use at any other location; nor may you enable operation of concurrent “instances” of the Software beyond the Authorized Scope of the Corporate license purchased. Corporate License Numbers are Confidential Information, and must be protected at a minimum to the extend set out in Section 8 of this Agreement.

- Any Software provided under an “Evaluation” or “Trial” license is provided for your internal evaluation purposes only, pending your purchase of a commercial-use Software license and payment of the applicable license fees. Evaluation licenses may automatically be converted into commercial licenses (e.g., with Authorized Scope converted into Static, Network, or Dynamic license) upon payment of the applicable license fees to Licensor or issuance of the applicable license key (whichever occurs first). Upon such conversion, this Agreement shall continue in full force and effect, subject to the restrictions applicable to the new Authorized Scope.

- Any Software provided under a “Demonstration” license is provided for your internal demonstration purposes only, and not for commercial or research purposes. You may not use the Software provided under a Demonstration license to operate upon data other than the data provided to you under the Demonstration license. You may not use any other software in conjunction with the Software provided under a Demonstration license to operate upon data other than the data provided to you under the Demonstration license.

- If you have purchased a “Named User” license, then you may allow up to the number of named users for which you have purchased a license to operate the Software. A named user is defined as one individual user identified with a unique user name and password. Each named user may operate no more than one session from a single workstation at a time. The workstations must all be located at the designated site. Only named users may use the Software.

3. **Restrictions on Use.** You acknowledge that you are receiving LICENSED RIGHTS only. The Software may only be used internally, by you or your Authorized Users, with the license key provided, for your copy(ies) of the Software. If any Software is provided on separate media (e.g., a CD-ROM), you may make a reasonable number of copies solely for your internal backup purposes. You shall not directly or indirectly: (i) sell, rent, lease, redistribute or transfer any of the Software or any rights in any of the Software, including through “charge back” or any other reselling or redistribution within your organization of any usage capacity you have licensed, without the prior express written approval of Licensor, (ii) modify, translate, reverse engineer (except to the limited extent permitted by law), decompile, disassemble, create, derivative works based on, sublicense, or distribute any of the Software, (iii) use any Software for the benefit of any third parties (e.g., in an ASP, outsourcing or service bureau relationship), or in any way other than in its intended manner, (iv) remove any proprietary notice, labels, or marks on or in Software, or (v) disable or circumvent
any access control or related device, process or procedure established with respect to the Software, including the license key or any other part thereof. You are responsible for all use of the Software and for compliance with this Agreement; any breach by you or any user of the Software provided to you shall be deemed to have been incurred by you. Licensor reserves all rights not expressly granted. You agree and acknowledge that you have been provided sufficient information such that you do not need to reverse engineer the Software in any way to permit other products to interoperate with the Software. Sharing of accounts does not reduce the type or number of licenses you must purchase. For example, if you pool ten (10) users through a single account, you must have a license authorizing use by at least ten (10) users. In addition, “Multiplexing” or “pooling” of any Software or hardware (including through use of third party software products such as those made by Citrix Systems, Inc., or use of any terminal applications/emulators) does not reduce the type or number of licenses you must purchase; the required number of licenses you have must be greater than or equal to the number of distinct “inputs” to the multiplexing software or hardware “front end”.

4. NO SEPARATION OF COMPONENTS. The Software is licensed as a single product. Some Licensor software products combine separately available components into a single product (e.g., a software suite product may be comprised of multiple component products). When licensed as a combination product, the component parts may not be separated for use independently of the combination product. You must first purchase a license to each component of the combination product before you may use it independently of the combination product.

5. CLASSROOM LICENSES. Certain Software may be licensed under Classroom License programs. Software licensed under such programs is subject to the additional restrictions in this Section. Software licensed under a “Classroom License” is restricted to use by students and teachers only in support of the classroom instruction of students and related class-work, such as the preparation of classroom materials, and the review and grading of exams, papers and class-related projects. Students may use the Software for class-related laboratory work and homework assignments. Use of Software licensed under such a program for any purposes other than as described in this Section, including for commercial or production purposes, or for non-classroom related research, is strictly prohibited.

6. ADDITIONAL PURCHASES. Purchase of additional or changed licenses or Authorized Scope is subject to availability and current pricing. Licensor may, from time to time, update the available Authorized Scope plans, and add or delete from available plans. To the extent you purchase an upgrade to your license by expanding the Authorized Scope under a changed or additional plan, Licensor will provide you with the additional license terms and conditions governing your use of Software under such plans; all other terms and conditions of this Agreement shall remain in effect.

7. TERM AND TERMINATION. Unless otherwise agreed, the term (“Term”) of this Agreement shall continue until it is terminated; however, if you are receiving an Evaluation or Demonstration license, the Term shall expire thirty (30) days following the date you receive the Software. Licensor may terminate this Agreement (or any License) upon fifteen (15) days prior written notice if you breach this Agreement and do not cure the breach within fifteen (15) days following receipt of written notice of breach. Immediately upon any termination or expiration of this Agreement, you agree to: (i) pay all amounts owed to Licensor; (ii) cease use of the Software for which your rights have been terminated; (iii) upon request, return to Licensor (or destroy) all copies of the Software and any other Confidential Information or proprietary materials in your possession for which your rights have been terminated; and (iv) upon request, certify in writing your compliance with (ii) and (iii), above.

8. CONFIDENTIALITY. You agree to protect Licensor’s Confidential Information with the same degree of care used to protect your own Confidential Information (but in no event less than a reasonable standard of care), and not to use or disclose any portion of such Confidential Information to third parties, except as expressly authorized in this Agreement. You acknowledge that the Software, including its structure, organization and design constitute proprietary and valuable trade secrets (and other intellectual property rights) of Licensor and/or its licensors. The term “Confidential Information” means, collectively, non-public information that Licensor (and its licensors) provide and reasonably consider to be of a confidential, proprietary or trade secret nature, including but not limited to confidential elements of the Software and Licensor’s (and its licensors’) technology and know-how, whether in tangible or intangible form, and whether or not stored, compiled or memorialized physically, electronically, graphically, photographically, or in writing.
Confidential Information does not include any information which: (i) is, as of the time of its disclosure or thereafter becomes part of the public domain through no fault of the receiving party; (ii) can be demonstrated by credible evidence: (x) as rightfully known to the receiving party prior to the time of its disclosure, or (y) to have been independently developed by the receiving party; (iii) is subsequently learned from a third party not under a confidentiality obligation; or (iv) is required to be disclosed pursuant to a duly authorized subpoena, court order, or government authority, provided that the receiving party has provided prompt written notice and assistance to Licensor prior to such disclosure, so that Licensor may seek a protective order or other appropriate remedy to protect against disclosure.

9. **WARRANTY AND DISCLAIMER.** Licensor warrants that, for a period of thirty (30) days from the date of initial delivery of Software to your designated shipping address (as indicated in the applicable invoice), the Software will, under normal use, substantially perform the functions described in its accompanying documentation. THE FOREGOING EXPRESS WARRANTY REPLACES AND IS IN LIEU OF ALL OTHER WARRANTIES OR CONDITIONS BY LICENSOR, WHETHER EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO ANY IMPLIED OR OTHER WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, NON-INFRINGEMENT OR NON-MISAPPROPRIATION OF INTELLECTUAL PROPERTY RIGHTS OF A THIRD PARTY, CUSTOM, TRADE, QUIET ENJOYMENT, ACCURACY OF INFORMATIONAL CONTENT, OR SYSTEM INTEGRATION. NO WARRANTY IS MADE THAT ANY SOFTWARE WILL OPERATE IN AN ERROR FREE, UNINTERRUPTED OR COMPLETELY SECURE MANNER, OR IN COMBINATION WITH THIRD PARTY HARDWARE OR SOFTWARE PRODUCTS. YOU ACKNOWLEDGE THAT LICENSOR HAS NO CONTROL OVER THE SPECIFIC CONDITIONS UNDER WHICH YOU USE THE SOFTWARE. ACCORDINGLY, EXCEPT FOR THE FOREGOING EXPRESS WARRANTY, LICENSOR CANNOT AND DOES NOT WARRANT THE PERFORMANCE OF THE SOFTWARE OR ANY PARTICULAR RESULTS THAT MAY BE OBTAINED BY THE USE OF THE SOFTWARE. THE SOFTWARE DOES NOT REPLACE YOUR OBLIGATION TO EXERCISE YOUR INDEPENDENT JUDGMENT IN USING THE SOFTWARE. The warranty made by Licensor may be voided by abuse or misuse.

10. **EXCLUSIVE REMEDY.** Licensor’s sole obligation and liability, and your sole and exclusive remedy under the warranty set forth in Section 9 shall be for Licensor to use reasonable efforts to have the problem remedied, or have the defective product replaced, provided that Licensor is notified in writing of all warranty problems during the applicable warranty period.

11. **THIRD PARTY SOFTWARE AND DATABASES.** You understand that third party products integrated into the Software or provided for use with the Software may be subject to additional terms and conditions and/or license agreements from the applicable third party vendor, which shall govern over conflicting terms of this Agreement for purposes of your relationship with the third party vendor. You agree not to use any such third party product on a stand-alone basis independent of the Software, unless you have purchased the appropriate license from the third party vendor for use of such products. NOTWITHSTANDING ANYTHING TO THE CONTRARY IN THIS AGREEMENT, ALL THIRD PARTY SOFTWARE, DATABASES AND OTHER PROGRAMS AND SOFTWARE COMPONENTS ARE PROVIDED “AS IS” WITHOUT ANY WARRANTY WHATSOEVER FROM LICENSOR. ANY DATABASES OR OTHER INFORMATION PROVIDED BY LICENSOR ARE DESIGNED TO SUPPLEMENT OTHER SOURCES OF INFORMATION, AND ARE NOT INTENDED TO REPLACE YOUR PROFESSIONAL DISCRETION AND JUDGMENT. Licensor agrees, upon request and as Licensor’s sole liability and obligation, to have passed through to you (to the extent it may reasonably do so) any warranties and indemnifications provided by the applicable third party vendor of any third party products provided to you. To the extent any problem or liability arises from a third party product, you agree to seek recourse solely from the applicable third party vendor and not Licensor.

12. **LICENSOR INDEMNIFICATION.** Licensor agrees to indemnify and defend you against any claims, actions, suits and proceedings by unaffiliated third parties arising from or related to a claim that the Software (other than any third party components or elements) infringes upon such party’s copyright, and Licensor agrees to pay all costs, losses, damages, and attorneys’ fees that a court finally awards, and all associated settlements; provided that, Licensor receives (i) prompt written notice of the claim; (ii) all necessary assistance, information and authority necessary for Licensor to defend the claim and perform Licensor’s obligations under this Section; and (iii) sole control of the defense of such claim and all associated negotiations. If such a claim is made or appears likely to be made, you agree to permit Licensor to enable you to continue to use the affected materials, or to have them modified them to make them non-infringing, or to have them
replaced with a substantially functional equivalent. If Licensor determines that none of these options is reasonably available, then Licensor may terminate this Agreement in whole or with respect to the affected product, and you may be entitled to a credit equal to the price paid for the affected product, less depreciation. THIS SECTION STATES LICENSOR’S AND ITS AFFILIATES’ ENTIRE OBLIGATION AND LIABILITY REGARDING INFRINGEMENT OR CLAIMS OF INFRINGEMENT. Licensor will have no responsibility for (x) any use of any product after you have been notified to discontinue use, (y) alteration or combination of any product with third party materials, or any misuse or unauthorized use, or (z) failure to use provided updated or modified products to avoid a claim of infringement or misappropriation.

13. **INDEMNIFICATION BY YOU.** You agree to indemnify and defend Licensor, its licensors, and its affiliates, against any third party claims arising from or related to your use or misuse of the Software or any breach of the terms and conditions of this Agreement, and you agree to pay all costs, losses, damages, and attorneys’ fees that a court finally awards, and all associated settlements.

14. **LIMITATION OF LIABILITY.** EXCEPT TO THE EXTENT PROHIBITED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR’S OR ITS AFFILIATES’ TOTAL, AGGREGATE LIABILITY ARISING FROM OR RELATED TO THIS AGREEMENT OR THE SOFTWARE (INCLUDING FOR NEGLIGENCE, STRICT LIABILITY, BREACH OF CONTRACT, MISREPRESENTATION, AND OTHER CONTRACT OR TORT CLAIMS), EXCEED THE AMOUNT OF YOUR DIRECT DAMAGES ACTUALLY INCURRED, UP TO THE AMOUNT OF FEES PAID TO LICENSOR UNDER THIS AGREEMENT FOR THE SOFTWARE THAT IS THE SUBJECT OF THE CLAIM UNDERLYING THE DAMAGES.

15. **EXCLUSION OF OTHER DAMAGES.** EXCEPT TO THE EXTENT PROHIBITED BY APPLICABLE LAW, UNDER NO CIRCUMSTANCES SHALL LICENSOR, ITS AFFILIATES, OR ANY OF THEIR SUPPLIERS OR LICENSORS BE LIABLE FOR ANY OF THE FOLLOWING: (I) THIRD PARTY CLAIMS, EXCEPT AS PROVIDED IN SECTION 12, (II) LOSS OR DAMAGE TO ANY SYSTEMS, RECORDS OR DATA, OR (III) INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL, PUNITIVE, RELIANCE, OR COVER DAMAGES (INCLUDING FOR LOST PROFITS, LOST SAVINGS AND DAMAGE TO ANY DATA OR SYSTEMS); EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. YOU ARE SOLELY RESPONSIBLE AND LIABLE FOR VERIFYING THE ACCURACY AND ADEQUACY OF ANY OUTPUT FROM THE SOFTWARE, AND FOR ANY RELIANCE THEREON.

16. **VERIFICATION.** Licensor shall have the right to have on-site audits periodically conducted of your use of the Software. These audits will generally be conducted during regular business hours, and Licensor will use its reasonable efforts not to interfere unduly with your regular business activities. Licensor may also require you to accurately complete a self-audit questionnaire in a form Licensor may have provided. If an audit reveals unauthorized use, you must promptly order sufficient licenses and Authorized Scope to permit all usage disclosed. If material unlicensed use is found (i.e., license shortage of 5% or more), you also shall reimburse Licensor for its costs incurred in connection with the verification.

17. **LEGAL COMPLIANCE; RESTRICTED RIGHTS.** Software is provided solely for lawful purposes and use. You shall be solely responsible for, and agree to comply with, all applicable laws, statutes, ordinances, and other governmental authority, however designated. Without limiting the foregoing, you agree to comply with all U.S. export laws and applicable export laws of your locality (if you are not in the United States), and shall not export any Software or other materials provided by Licensor without first obtaining all required authorizations or licenses. Products acquired for use within or for any United States federal agency are provided with “LIMITED RIGHTS” and “RESTRICTED RIGHTS” as defined in DFARS 252.227-7013 and FAR 52.227-19.

18. **GOVERNING LAW; SEVERABILITY.** This Agreement shall be governed in all respects by the laws of the State of California, USA, without regard to choice-of-law rules or principles. Any dispute arising out of or related to this Agreement shall be resolved only in the state or federal courts having jurisdiction in California. If any provision of this Agreement is held to be illegal or unenforceable for any reason, then such provision shall be deemed to be restated so as to be enforceable to the maximum extent permissible under law; the remainder of this Agreement shall remain in full force and effect.

19. **GENERAL.** This Agreement, including any associated Licensor invoices (all of which are incorporated herein), are collectively the parties’ complete agreement regarding its subject matter, superseding any prior oral or written communications. You understand and agree that, to the extent Licensor permits you to
use a non-Licensor purchase order or other form to order Software, Licensor does so solely for your convenience. Any terms in any such forms that purport to vary or are in addition to any terms in this Agreement or in the applicable Licensor form covering your purchase shall be deemed to be void and of no effect. Amendments or changes to this Agreement must be in mutually executed writings to be effective. Sections 1, 3–8, 10–19, and all warranty disclaimers and use restrictions shall survive the termination or expiration of this Agreement. The parties are independent contractors for all purposes under this Agreement. Neither party shall be liable for any delay or failure due to force majeure and other causes beyond its reasonable control; this provision shall not apply to any of your payment obligations. Any notices under this Agreement to Licensor must be personally delivered or sent by certified or registered mail, return receipt requested, or by nationally recognized overnight express courier, to the address specified herein or such other address as Licensor may specify in writing. Such notices will be effective upon receipt, which may be shown by confirmation of delivery. All such notices shall be sent to the attention of General Counsel of Invitrogen Corporation (unless otherwise specified by Licensor). You may not assign or otherwise transfer this Agreement without Licensor’s prior written consent. This Agreement shall be binding upon and inure to the benefit of the parties’ successors and permitted assigns. You agree, at Licensor’s request and reasonable expense, to provide reasonable assistance and cooperation to Licensor and its designees, and to give testimony and execute documents and to take such further acts reasonably requested by the other to acquire, transfer, maintain, perfect, and enforce Licensor’s intellectual property rights as described in this Agreement. To the extent you fail to do so, you appoint Licensor’s or its affiliates’ officers as your attorney in fact to execute documents on your (and your personnel’s), successors’ and assigns’ behalf for this limited purpose.
II. Software Product License Agreement Terms for Academic and/or Not-For-Profit Users

This is a legal agreement (“Agreement”) between you, the academic or not-for-profit researcher receiving a product license, and Invitrogen Corporation, having offices at 1600 Faraday Avenue, Carlsbad California 92008 USA (“Licensor”). This Agreement is part of a package that includes one or more software products and certain electronic and/or written materials. This Agreement covers your licensing of the Software (as the term is defined below). BY CLICKING YOUR ACCEPTANCE OF THIS AGREEMENT, OR BY INSTALLING OR USING THE SOFTWARE OR ANY OTHER COMPONENT OF THE PACKAGE, YOU ACKNOWLEDGE AND REPRESENT THAT:

(1) YOU ARE AN ACADEMIC OR NOT-FOR-PROFIT RESEARCHER INTENDING TO USE THE SOFTWARE ONLY FOR NOT-FOR-PROFIT AND/OR ACADEMIC RESEARCH PURPOSES; and

(2) YOU HAVE READ ALL OF THE TERMS AND CONDITIONS OF THIS AGREEMENT, UNDERSTAND THEM, AND AGREE TO BE LEGALLY BOUND BY THEM.

If you do not agree to the terms of this Agreement, you may not install or use the Software.

“Designated Site” means your facilities or offices located at the postal address provided by you, unless otherwise indicated in a license key provided to you; and “Software” means the software product(s) accompanying this Agreement, including the associated user manuals, user documentation and application program interfaces, and license key(s) provided, and any patches, updates, upgrades, improvements, enhancements, fixes and revised versions of any of the foregoing that may be provided to you from time to time, and any combination of the foregoing.

1. TAXES; SHIPPING. If any authority imposes a duty, tax or similar amount on the transfer of Software to you (other than based on Licensor’s income), you agree to pay, or to promptly reimburse Licensor for, all such amounts. All shipments by Licensor or its designee are FCA point of shipment.

2. GRANT OF SOFTWARE LICENSE. Subject to the terms and conditions of this Agreement, Licensor grants to you a non-exclusive, non-transferable license (the “Academic Vector NTI License”) to use the Software for your internal operations and internal data processing purposes within and up to the Authorized Scope for which you have registered the Software for use. The term “Authorized Scope” means the following, and any other capacity or use restrictions on the license granted to you: You have received a “Personal Static” license. For each such license you may install and operate the Software on a designated computer located at the Designated Site. You may not network the Software for use at any other location. Use of the Software is limited to the computer designated for each Static license.

3. RESTRICTIONS ON USE. You acknowledge that you are receiving LICENSED RIGHTS only. The Software may only be used internally, by you, with the license key provided, for your copy(ies) of the Software. If any Software is provided on separate media (e.g., a CD-ROM), you may make a reasonable number of copies solely for your internal backup purposes. You shall not directly or indirectly: (i) sell, rent, lease, redistribute or transfer any of the Software or any rights in any of the Software, including through “charge back” or any other reselling or redistribution within your organization of any usage capacity you have licensed, without the prior express written approval of Licensor, (ii) modify, translate, reverse engineer (except to the limited extent permitted by law), decompile, disassemble, create, derivative works based on, sublicense, or distribute any of the Software, (iii) use any Software for the benefit of any third parties (e.g., in an ASP, outsourcing or service bureau relationship), or in any way other than in its intended manner, (iv) remove any proprietary notice, labels, or marks on or in Software, or (v) disable or circumvent any access control or related device, process or procedure established with respect to the Software, including the license key or any other part thereof. You are responsible for all use of the Software and for compliance with this Agreement; any breach by you or any user of the Software provided to you shall be deemed to have been incurred by you. Licensor reserves all rights not expressly granted. You agree and acknowledge that you have been provided sufficient information such that you do not need to reverse engineer the Software in any way to permit other products to interoperate with the Software.

4. NO SEPARATION OF COMPONENTS. The Software is licensed as a single product. Some Licensor software products combine separately available components into a single product (e.g., a software suite product may be comprised of multiple component products). When licensed as a combination product, the component parts may not be separated for use independently of the combination product. You must first pur-
chase a license to each component of the combination product before you may use it independently of the combination product.

5. **TERM AND TERMINATION.** Unless otherwise agreed, the term ("Term") of this Agreement shall continue until it is terminated. Licensor may terminate this Agreement (or any License) upon fifteen (15) days prior written notice if you breach this Agreement and do not cure the breach within fifteen (15) days following receipt of written notice of breach. Immediately upon any termination or expiration of this Agreement, you agree to: (i) cease use of the Software for which your rights have been terminated; (ii) upon request, return to Licensor (or destroy) all copies of the Software and any other Confidential Information or proprietary materials in your possession for which your rights have been terminated; and (iii) upon request, certify in writing your compliance with (i) and (ii), above.

6. **CONFIDENTIALITY.** You agree to protect Licensor's Confidential Information with the same degree of care used to protect your own Confidential Information (but in no event less than a reasonable standard of care), and not to use or disclose any portion of such Confidential Information to third parties, except as expressly authorized in this Agreement. You acknowledge that the Software, including its structure, organization and design constitute proprietary and valuable trade secrets (and other intellectual property rights) of Licensor and/or its licensors. The term "Confidential Information" means, collectively, non-public information that Licensor (and its licensors) provide and reasonably consider to be of a confidential, proprietary or trade secret nature, including but not limited to confidential elements of the Software and Licensor's (and its licensors') technology and know-how, whether in tangible or intangible form, and whether or not stored, compiled or memorialized physically, electronically, graphically, photographically, or in writing. Confidential Information does not include any information which: (i) is, as of the time of its disclosure or thereafter becomes part of the public domain through no fault of the receiving party; (ii) can be demonstrated by credible evidence: (x) as rightfully known to the receiving party prior to the time of its disclosure, or (y) to have been independently developed by the receiving party; (iii) is subsequently learned from a third party not under a confidentiality obligation; or (iv) is required to be disclosed pursuant to a duly authorized subpoena, court order, or government authority, provided that the receiving party has provided prompt written notice and assistance to Licensor prior to such disclosure, so that Licensor may seek a protective order or other appropriate remedy to protect against disclosure.

7. **WARRANTY AND DISCLAIMER.** Your use of the Software and its results is entirely at your own risk. Should the Software prove defective, you assume the entire cost of all service, repair or correction. The Software is made available to you "AS IS", AND LICENSOR MAKES NO REPRESENTATION OR WARRANTY REGARDING THE SOFTWARE, AND HEREBY DISCLAIMS ALL WARRANTIES, WHETHER EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO ANY IMPLIED OR OTHER WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE, NON-INFRINGEMENT OR NON-MISAPPROPRIATION OF INTELLECTUAL PROPERTY RIGHTS OF A THIRD PARTY, CUSTOM, TRADE, QUIET ENJOYMENT, ACCURACY OF INFORMATIONAL CONTENT, OR SYSTEM INTEGRATION. NO WARRANTY IS MADE THAT ANY SOFTWARE WILL OPERATE IN AN ERROR FREE, UNINTERRUPTED OR COMPLETELY SECURE MANNER, OR IN COMBINATION WITH THIRD PARTY HARDWARE OR SOFTWARE PRODUCTS. YOU ACKNOWLEDGE THAT LICENSOR HAS NO CONTROL OVER THE SPECIFIC CONDITIONS UNDER WHICH YOU USE THE SOFTWARE. ACCORDINGLY, EXCEPT FOR THE FOREGOING EXPRESS WARRANTY, LICENSOR CANNOT AND DOES NOT WARRANT THE PERFORMANCE OF THE SOFTWARE OR ANY PARTICULAR RESULTS THAT MAY BE OBTAINED BY THE USE OF THE SOFTWARE. THE SOFTWARE DOES NOT REPLACE YOUR OBLIGATION TO EXERCISE YOUR INDEPENDENT JUDGMENT IN USING THE SOFTWARE. The warranty made by Licensor may be voided by abuse or misuse. Licensor has no obligation to provide any support, error corrections, updates, or any other services related to the Software ("Services"). ANY SERVICES RELATED TO THE SOFTWARE ARE PROVIDED "AS IS", AND LICENSOR MAKES NO REPRESENTATION OR WARRANTY REGARDING SUCH SERVICES, AND HEREBY DISCLAIMS ALL WARRANTIES, WHETHER EXPRESS, IMPLIED, OR STATUTORY, WITH RESPECT TO SUCH SERVICES.

8. **THIRD PARTY SOFTWARE AND DATABASES.** You understand that third party products integrated into the Software or provided for use with the Software may be subject to additional terms and conditions and/or license agreements from the applicable third party vendor, which shall govern over conflicting terms of this Agreement for purposes of your relationship with the third party vendor. You agree not to use any such
third party product on a stand-alone basis independent of the Software, unless you have purchased the appropriate license from the third party vendor for use of such products. NOTWITHSTANDING ANYTHING TO THE CONTRARY IN THIS AGREEMENT, ALL THIRD PARTY SOFTWARE, DATABASES AND OTHER PROGRAMS AND SOFTWARE COMPONENTS ARE PROVIDED "AS IS" WITHOUT ANY WARRANTY WHATSOEVER FROM LICENSOR. ANY DATABASES OR OTHER INFORMATION PROVIDED BY LICENSOR ARE DESIGNED TO SUPPLEMENT OTHER SOURCES OF INFORMATION, AND ARE NOT INTENDED TO REPLACE YOUR PROFESSIONAL DISCRETION AND JUDGMENT. Licensor agrees, upon request and as Licensor’s sole liability and obligation, to have passed through to you (to the extent it may reasonably do so) any warranties and indemnifications provided by the applicable third party vendor of any third party products provided to you. To the extent any problem or liability arises from a third party product, you agree to seek recourse solely from the applicable third party vendor and not Licensor.

9. **Indemnification by You.** You agree to indemnify and defend Licensor, its licensors, and its affiliates, against any third party claims arising from or related to your use of the Software or any breach of the terms and conditions of this Agreement, and you agree to pay all costs, losses, damages, and attorneys’ fees that a court finally awards, and all associated settlements.

10. **Exclusion of Other Damages.** Except to the extent prohibited by applicable law, under no circumstances shall Licensor, its affiliates, or any of their suppliers or licensors be liable for any of the following: (i) third party claims (ii) loss or damage to any systems, records or data, or (iii) direct, indirect, incidental, special, consequential, punitive, reliance, or cover damages (including for lost profits, lost savings and damage to any data or systems); even if advised of the possibility of such damages. You are solely responsible and liable for verifying the accuracy and adequacy of any output from the Software, and for any reliance thereon.

11. **Limitation of Liability.** Except to the extent prohibited by applicable law, in no event will Licensor’s or its affiliates’ total, aggregate liability arising from or related to this Agreement or the Software (including for negligence, strict liability, breach of contract, misrepresentation, and other contract or tort claims) exceed the amount of fees paid by you to Licensor under this Agreement for the Software that is the subject of the claim underlying the damages, or one hundred dollars ($100.00), whichever is less.

12. **Verification.** Licensor shall have the right to have on-site audits periodically conducted of your use of the Software. These audits will generally be conducted during regular business hours, and Licensor will use its reasonable efforts not to interfere unduly with your regular business activities. Licensor may also require you to accurately complete a self-audit questionnaire in a form Licensor may have provided. If an audit reveals unauthorized use, you must promptly order sufficient licenses and Authorized Scope to permit all usage disclosed. If material unlicensed use is found (i.e., license shortage of 5% or more), you also shall reimburse Licensor for its costs incurred in connection with the verification.

13. **Legal Compliance; Restricted Rights.** Software is provided solely for lawful purposes and use. You shall be solely responsible for, and agree to comply with, all applicable laws, statutes, ordinances, and other governmental authority, however designated. Without limiting the foregoing, you agree to comply with all U.S. export laws and applicable export laws of your locality (if you are not in the United States), and shall not export any Software or other materials provided by Licensor without first obtaining all required authorizations or licenses. Products acquired for use within or for any United States federal agency are provided with “LIMITED RIGHTS” and “RESTRICTED RIGHTS” as defined in DFARS 252.227-7013 and FAR 52.227-19.

14. **Governing Law; Severability.** This Agreement shall be governed in all respects by the laws of the State of California, USA, without regard to choice-of-law rules or principles. Any dispute arising out of or related to this Agreement shall be resolved only in the state or federal courts having jurisdiction in California. If any provision of this Agreement is held to be illegal or unenforceable for any reason, then such provision shall be deemed to be restated so as to be enforceable to the maximum extent permissible under law; the remainder of this Agreement shall remain in full force and effect.
15. **GENERAL.** This Agreement is the parties’ complete agreement regarding its subject matter, superseding any prior oral or written communications. You understand and agree that, to the extent Licensor permits you to use a non-Licensor purchase order or other form to order Software, Licensor does so solely for your convenience. Any terms in any such forms that purport to vary or are in addition to any terms in this Agreement or in the applicable Licensor form covering your purchase shall be deemed to be void and of no effect. Amendments or changes to this Agreement must be in mutually executed writings to be effective. Sections 1, 3–8, 11–15, and all warranty disclaimers and use restrictions shall survive the termination or expiration of this Agreement. The parties are independent contractors for all purposes under this Agreement. Neither party shall be liable for any delay or failure due to force majeure and other causes beyond its reasonable control. Any notices under this Agreement to Licensor must be personally delivered or sent by certified or registered mail, return receipt requested, or by nationally recognized overnight express courier, to the address specified herein or such other address as Licensor may specify in writing. Such notices will be effective upon receipt, which may be shown by confirmation of delivery. All such notices shall be sent to the attention of General Counsel of Invitrogen Corporation (unless otherwise specified by Licensor). You may not assign or otherwise transfer this Agreement without Licensor’s prior written consent. This Agreement shall be binding upon and inure to the benefit of the parties’ successors and permitted assigns. You agree, at Licensor’s request and reasonable expense, to provide reasonable assistance and cooperation to Licensor and its designees, and to give testimony and execute documents and to take such further acts reasonably requested by the other to acquire, transfer, maintain, perfect, and enforce Licensor’s intellectual property rights as described in this Agreement. To the extent you fail to do so, you appoint Licensor’s or its affiliates’ officers as your attorney in fact to execute documents on your (and your personnel's), successors’ and assigns’ behalf for this limited purpose.

I HAVE READ AND AGREE TO BE BOUND BY THE ABOVE TERMS THAT APPLY TO ME AS EITHER A COMMERCIAL USER (SECTION I) OR AN ACADEMIC USER (SECTION II), AS APPLICABLE.
Welcome to \textit{Vector NTI Advance}, part of a family of software packages developed by Invitrogen, Carlsbad, California. Other life science applications developed by Invitrogen include \textit{Vector Xpression}, \textit{Vector PathBlazer}, \textit{LabShare} for Vector NTI\textsuperscript{®} and \textit{Vector NTI}\textsuperscript{®} for Mac OS X.

Vector NTI Advance is made up of the following modules:

- Vector NTI\textsuperscript{™}
- AlignX\textsuperscript{®}
- BioAnnotator\textsuperscript{™}
- ContigExpress\textsuperscript{®}
- GenomBench\textsuperscript{™}

You may not have purchased licenses for all of the modules of Vector NTI Advance. If you would like to do so, please contact Invitrogen at their website: \url{www.invitrogen.com/bioinformatics} for more information.

Getting Started with Vector NTI Advance

- To learn about the \textit{Vector NTI Advance 10 User’s Manual} structure, review Chapter 1.
- To read a brief overview of the Vector NTI Advance software, review Chapter 2.
- To activate your license for Vector NTI Advance 10:
  - Refer to the \textit{Vector NTI Advance 10 Installation and Licensing Guide} that you received when purchasing Vector NTI Advance 10
  - Download the \textit{Vector NTI Advance 10 Installation and Licensing Guide} from the Invitrogen website, \url{www.invitrogen.com/bioinformatics}. 
or

- See the section on License Manager in Chapter 39, page 650 in this user’s manual.
- To start Vector NTI Advance, refer to Launching Vector NTI on page 13.
- To learn various methods of opening and using of Online Help, refer to Online Help on page 13.
- To learn independently how to use components of the Vector NTI software, walk through the tutorials, Chapters 4 through 16.
- To learn about the global features used in all Vector NTI Advance module viewers, periodically review Chapter 3.
- To learn more about the automation features of Vector NTI, refer to the Invitrogen website: www.invitrogen.com/bioinformatics.

The Vector NTI User

As Vector NTI Advance uses a convenient graphical user interface (GUI), in using the software you are expected to have some familiarity with window environments and web browsers.

The Vector Advance 10 User’s Manual Format

To be a useful and simple-to-use resource, the Vector NTI Advance User’s Manual contains:

- Section I: Three introductory chapters, 1 through 3, covering Vector NTI basics.
- Section II: Vector NTI Advance Tutorial chapters, chapters 4 through 16, enabling VNTI users at any level to learn independently how to use components of the Vector NTI software.
- Section III: Reference tools, chapters 17 through 27 organized around the functions and features of Vector NTI.
- Section IV: Vector NTI Advance Applications, chapters 28 through 39, covering additional components of Vector NTI Advance.
- Section VI: Appendices
- Section VII: A complete index

Use the manual index to easily look up and find descriptions of specific features and functions. Electronic versions of the Vector NTI Advance 10 User’s Manual and the Vector NTI Advance 10 Installation and Licensing Guide are also available at the Invitrogen website: www.invitrogen.com/bioinformatics.

Conventions Used in this Manual

The following table uses various typefaces to differentiate between regular text and menu commands, keyboard keys, toolbar buttons, dialog box options and text that you type. The table describes how conventions are represented in this manual.

<table>
<thead>
<tr>
<th>Convention</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bold and Capitalized Command</td>
<td>Menu command</td>
<td>Database &gt; Explore</td>
</tr>
<tr>
<td>Capitalized command &gt; Capitalized command</td>
<td>Sequential Menu commands</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1.1 User Manual Conventions*
<table>
<thead>
<tr>
<th><strong>Convention</strong></th>
<th><strong>Description</strong></th>
<th><strong>Example</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>TEXT IN SMALL CAPS</td>
<td>Keyboard key that you press</td>
<td>The Enter Key ENTER</td>
</tr>
<tr>
<td>TEXT IN SMALL CAPS + TEXT IN SMALL CAPS</td>
<td>Keys that you press together</td>
<td>Press SHIFT+ CTRL, then release both.</td>
</tr>
<tr>
<td>TEXT IN SMALL CAPS followed by TEXT IN SMALL CAPS</td>
<td>Keys that you press in sequence</td>
<td>Press ENTER, followed by ENTER</td>
</tr>
<tr>
<td>Icon</td>
<td>Small picture on a toolbar button that you press</td>
<td>Click the <strong>Zoom In</strong> button (🔍) to enlarge the image</td>
</tr>
<tr>
<td><strong>Boldface type</strong></td>
<td>Options or tabs that you select in dialog boxes, list boxes, or drop-down menus Buttons that you click</td>
<td>In the Restriction Enzymes dialog box, select the enzyme(s) and press the <strong>Add</strong> button.</td>
</tr>
<tr>
<td><strong>Italic boldface type</strong></td>
<td>Text that you type</td>
<td>In the New Molecule dialog box, enter the molecule name, <em>my pBR322</em>.</td>
</tr>
<tr>
<td><strong>Note:</strong></td>
<td>Highlights a particular concept or extends the current topic</td>
<td><strong>Note:</strong> To calculate Separation Time, at least two fragments must be selected.</td>
</tr>
<tr>
<td><strong>Important:</strong> Or <strong>Warning!</strong></td>
<td>Highlights information of which you should be particularly aware</td>
<td><strong>Important:</strong> Once parent-descendant relationships have been disconnected, they CANNOT be reconnected.</td>
</tr>
</tbody>
</table>

*Table 1.1 User Manual Conventions (continued)*
Introduction

Vector NTI Advance, a suite of desktop applications, is a comprehensive set of sequence analysis software tools specifically designed for the molecular biologist to study, visualize, manipulate, construct and store biological molecules. With this system, you can:

- Easily navigate and search the user-defined, password-protected molecule and enzyme databases
- Access molecules across a network with user-defined and password-protected databases
- Generate recombinant strategies and protocols, and design and analyze PCR primers and other oligos
- Animate virtual gels for restriction fragment analysis
- Analyze physiochemical properties of molecule sequences
- Assemble contigs from text sequences or chromatograms
- Create and customize publication-quality graphics
- Take advantage of full Internet connectivity with more than 20 Internet sites for searching and importing annotated molecules back into Vector NTI Advance
- Automate common tasks such as batch BLAST, batch annotation, batch primer design, etc. through the Application Programming Interfaces (APIs)

While all of the software components are fully capable of functioning independently of one another, the program integrates data, analyses and presentations easily among the applications.

Vector NTI Database Explorer

The Vector NTI database provides storage and organization functions for molecules and other objects used in cloning strategies, molecule manipulation and analysis, and sequence analysis. Database objects include DNA/RNA and protein molecules, restriction enzymes, oligonucle-
otides (including PCR primers, sequencing primers and hybridization probes), gel markers, BLAST search results, and citations.

In the Database Explorer window, you can:

- Create new objects
- Edit and delete old objects
- Perform database searches
- Organize objects into convenient groups (subsets)
- Import and export objects
- Create Vector NTI “archives” of objects to share with other users
- Format references according to bibliographic styles assigned to various professional publications

General database functions let you make backup copies of the database in the specified subfolder of your hard drive or floppy disk, restore all database files from a backup and compress the database by removing all unnecessary data from the data files and database tables.

Vector NTI also has two additional data managers:

- Contact Manager to create, modify, and organize the address data of persons or organizations with which you frequently communicate.
- User Field Manager to define data fields to describe your database objects.

Organizing your Biological Objects

Using the Vector NTI database, you can conveniently organize and effectively manage a large amount of molecular biology data. Database objects can be organized into an arbitrary number of subsets. The Database Explorer has convenient commands for creating and managing subsets of all database objects.

Description fields of database objects allow you to describe any object property in formal and semiformal ways. Vector NTI serves as a global “organizer” for your biological data.

Database Search

Vector NTI lets you search for database objects using many types of data as search conditions. These include, for example, form (circular, linear, Invitrogen), storage type (basic, constructed, translated), size, description/comment/user fields text, and so forth.

Import/Export Functions

You can import molecules (including their feature tables) from GenBank/GenPept, EMBL/SWISS-PROT and FASTA formats. You can also import nucleotide or amino acid sequences from an ASCII file of flexible format, and Vector NTI will automatically create the new database molecule and assign the sequence to the molecule.

Archives - A Tool For Sharing Database Information

Vector NTI lets you place your molecules, enzymes, oligonucleotides and gel markers into archives that can be transferred to another computer (Mac or PC) and read by Vector NTI. This allows you to share objects with colleagues, or to place them simultaneously on several computers. As molecules are archived, Vector NTI verifies the consistency of all information.
Creating New Molecules

There are five different ways of creating new DNA/RNA and protein molecules in Vector NTI:

1. Importing from GenBank/GenPept, EMBL/SWISS-PROT and FASTA formats, or from an ASCII file of flexible format. The sequence and Feature map are converted from the file, and the new molecule becomes part of the Vector NTI database.

2. Creating new molecules from user-defined nucleotide or amino acid sequences. These can be manually entered or pasted from the clipboard and the sequence entered as a new molecule.

3. Translating a coding region of an existing DNA or RNA molecule to create new protein molecules.

4. Construction of new DNA/RNA molecules from user-defined compatible component fragments from other molecules.

5. Design of new DNA/RNA molecules from components in a user-defined fragment list, using Vector NTI’s built-in biological knowledge to design the recombination process.

All new molecules are integrated into the database and can participate in all further operations and analyses.

Creating New Enzymes, Oligos, and Gel Markers

Enzyme, oligonucleotide, and gel marker objects can be created “from scratch” using Vector NTI’s editors or by importing them from a REBASE database.

Oligos and Primers

Vector NTI can design PCR primers, sequencing primers and hybridization probes and save them to the database for future use. Using parameters you have defined, Vector NTI can analyze those primers and probes (or those you have defined yourself) to determine the best ones for optimal experimental results, then link you to the Invitrogen™ website to order your custom primers.

Editing Database Objects

Full molecule editing capabilities are available in Vector NTI, either from Database Explorer or the Molecule Display window. In Molecule Display windows, you can edit molecules by inserting, deleting and replacing sequence fragments and features. You can also modify a molecule’s display format and general data. When your editing is complete, you can save the molecule into the database to make the changes permanent.

Parent-Descendant Connections

As you modify molecules, Vector NTI monitors and manages parent-descendant connections to ensure data consistency through all the descendants (see Appendix D, page 669).

Generating Restriction Maps

Restriction maps of DNA/RNA molecules can be quickly generated in Vector NTI. For unsequenced molecule regions, you may enter the known positions of restriction sites. All the molecule descendants inherit these sites.
Molecule Display

Molecule display windows are Vector NTI's convenient graphical user interface (GUI) for inspecting and editing text description, graphical maps and analyses, and sequences of DNA/RNA and protein molecules. They are also powerful and convenient graphical tools for selecting fragments for molecule construction and design and for preparing publication-quality images.

Molecule display windows consist of Text, Graphics and Sequence Panes.

Text Pane

Text Panes contain a general molecule description and folders describing a molecule’s features, search results, cloning strategies and protocols, and biological data.

Graphics Pane

A molecule display window’s Graphics Pane contains a graphically displayed or linear molecule map. In the Graphics Pane, you can visually select and “prepare” fragments for construction and design of new molecules, search for probes and primers, or for restriction maps.

All graphics objects are movable, resizable and editable.

Sequence Pane

A molecule display window’s Sequence Pane contains the formatted nucleotide or amino acid sequence of a molecule.

For DNA/RNA molecules, Vector NTI can show restriction sites, motifs, ORFs and functional signals, and translated reading frames.

For protein molecules, Vector NTI can display protein features together with the amino acid sequence.

You can print any molecule display window or you can save it to an external file or clipboard. From the clipboard, you can insert maps, sequences and molecule descriptions directly into word processing documents.

Creating Molecule Documents

You can save all the information from a molecule display window to permanent storage by creating a Molecule Document file. Since the Molecule Document format is based on GenBank/GenPept formats, not only Vector NTI, but any program able to import GenBank/GenPept files can also import Vector NTI Molecule Documents.

Gel Display Windows

Gel Display windows can be created for different types of electrophoresis. The Gel Display window’s Graphics Pane replicates the results of an actual gel.

Creating Gel Display Windows

To create a new Gel Display window, you must define its parameters as to electrophoresis and buffer types, electrophoresis conditions and display parameters.

You can modify the settings of a Gel Display window at any time after creation for better separation of the molecules’ fragments.

Creating Gel Samples and Gel Markers

In Vector NTI, a gel lane may contain either a Gel Sample or a Gel Marker, user-generated or from the database. Gel Samples are created by “cutting” database molecules with restriction
enzymes. Gel Markers, sets of fragments of known lengths, can be loaded either from the database or can be user-generated.

Running the Gel

After a Gel Display window is created, and your samples and markers have been “added” to the “lanes,” you can run a gel according to selected time increments or animation settings. You can color-code specific fragments or define “separation distances” or time lengths to better visualize fragment separation. Because of the simplicity of their generation, Gel Samples are created on the spot for short-term use, but are not stored in the Vector NTI database.

Align X

AlignX is a multiple sequence alignment tool. In AlignX, you can:

- Perform multiple sequence alignments on either proteins or nucleic acids without reformating between applications and computer platforms.
- In the alignment, link, localize and compare important domains in individual sequences
- Generate plots of similarity and sequence complexity

AlignX features include multicolor presentation, alignment editing, sequence comparison dot plots, guide trees, sequence and structure alignments, and similarity distribution plots. AlignX reads all standard text formats such as FASTA, GenBank, EMBL, SWISS-PROT, GenPept, and ASCII text.

BioAnnotator

BioAnnotator is a comprehensive set of protein and nucleic acid sequence analysis tools, offering over 50 different predefined protein scales linked with Feature maps and sequences. BioAnnotator also offers the following advanced search and analysis tools, which are run in the Analysis Monitor (see below): ProSite, PFAM, BLOCKS and Proteolytic cleavage.

BioAnnotator is formatted so that plots can be filed, overlapped, scaled, moved and formatted to fit your presentation needs. Sequence regions can be highlighted by changing font face, size, and color, and the plots linked to a molecule feature table. The use of a standard text file format in all programs eliminates the need to reformat data for use in any Vector NTI Advance modules.

ContigExpress

ContigExpress is a program for assembling many small fragments—either text sequences or chromatograms from automated sequencers—into longer, contiguous sequences. You can manage fragments and assemblies in an Explorer window, and open fragments and contigs in display windows for further study and manipulation. The multi-pane windows link sequences with their properties, chromatograms and graphic representations.

You can edit nucleic acid fragments directly with the chromatograms in full view. Changes are tracked and an edit history is maintained. The contigs generated are then saved and managed as GenBank, EMBL or FASTA files. No exporting or reformatting of proprietary file formats is necessary. Just drag and drop into Vector NTI to map and analyze.

GenomBench

GenomBench is Vector NTI Advance’s genome project workhorse. GenomBench can retrieve data from compatible DAS-compliant servers, such as UCSC and Ensembl. It provides an interface for browsing and querying such sites using keywords, flanking markers, and BLAT (UCSC
public database) or SSAHA (Ensembl) similarity searches. GenomBench also provides a means for mapping your data onto a genomic region using Sim4 or Spidey analyses.

As with other Vector NTI Advance applications, GenomBench has a multi-pane user interface where you can view important text information in the project, and navigate, manipulate and annotate a graphical representation of the genome. You can also perform sequence searches, translations and alignments. In addition, GenomBench provides tools for querying public databases, copying and pasting sequence and graphics, performing feature searches, importing and exporting annotated sequences and sending annotated sequence segments to Vector NTI.

Like Vector NTI, GenomBench has a Local database for storing your sequences and analysis results.

**Remote Services Tools**

Vector NTI provides applications that allow you to use remote bio-services for analyzing your data. You can store the results of such analyses in the Vector NTI Database and share them with your colleagues. Vector NTI also provides a tool to help you configure links to the remote servers used for analysis and data storage.

**BLAST Search/BLAST Viewer**

Vector NTI Advance provides you with a search engine for launching BLAST searches of the GenBank databases at the National Center for Biotechnology Information website. Results are listed in the BLAST Search dialog box from which they can be opened and reviewed in BLAST Viewer, displaying a linear representation of aligned hit elements from the hit molecules. BLAST results can be saved to the Vector NTI Database from the BLAST Viewer.

*Note:* BLAST searches can also be launched from Analysis Monitor, described below.

**PubMed/Entrez Search, Citation Viewer and Citation Table of Database Explorer**

These tools work in tandem to provide a practical and efficient way to search public databases for molecules or citations, and retrieve them, complete with molecule files or abstracts that can be viewed and stored.

**Remote Database Sources Wizard**

The Remote Database Sources Wizard steps you through the process of creating links to remote servers, which enable Vector NTI remote services tools to connect to external databases, such as Entrez and LabShare. Basic links for BLAST and PubMed/Entrez searches are already created; you do not have to create them.

**Miscellaneous Tools**

**Analysis Monitor**

The Vector NTI Analysis Monitor provides a convenient interface for performing the following advanced DNA/RNA and Protein molecule analyses included in BioAnnotator and GenomBench: PFAM, ProSite, Blocks and Proteolytic Cleavage, and Spidey and Sim4 analyses, respectively. Analysis Monitor also provides an alternative interface for launching BLAST searches.

Vector NTI functions as a viewer for analysis results retrieved by the Analysis Monitor. The Analysis Results data type enables you to store results from the Analysis Monitor application.

Analyses that are included with GenomBench or BioAnnotator licenses also become available from the Vector NTI window when these products are licensed.
GCG Converter

GCG Converter is a tool in Vector NTI Advance that converts sequences in GCG file formats so that they can be imported into the software.

Matrix Editor

Matrix Editor allows inspection of and editing of matrices. Matrices are used in two Vector NTI Advance applications, AlignX and AlignX Blocks. Matrices can be viewed in either program in the Alignment Setup dialog box, but editing can only take place in Matrix Editor.

AlignX Blocks

AlignX Blocks is a program for locating, analyzing and editing blocks of localized sequence similarity among multiple protein sequences and linking them into a composite multiple alignment.

3D-Mol

The 3D-Mol application of Vector NTI Advance allows you to visualize and manipulate 3-dimensional molecule structures described in a number of standard molecule structure files, including PDB (Protein Data Bank). 3D-Mol also supports creation of user-defined structures and provides tools for aligning protein chains and calculating protein surfaces. Structures can be saved to a 3D-Mol project and/or exported to PDB or M3D files, or to Vector NTI.

Oligo Files Importer

The Oligo File Importer utility allows you to import oligo lists into the Vector NTI database. The oligo list to be imported must be in one of three file formats for use with the Oligo utility: Comma delimited format (.csv format), Tab delimited format or Semicolon delimited format.

Tools Manager

Vector NTI is shipped with more than 20 tools providing connection to various Internet servers. You can create your own tools connecting Vector NTI to local or remote programs using the configuration mechanism described in this chapter. All of the Vector NTI tools are organized and managed using Tools Manager.

License Manager

Vector NTI supports three different types of licenses: static, dynamic and trial licenses. When unlicensed, Vector NTI operates in Demo Mode for the purpose of demonstrating the basic features of the software. The various license types were developed to meet users’ needs. For example, users in different industrial, scientific and educational environments require different sharing and allotment privileges. These privileges are all administered through the License Manager.

Vector NTI System Information

Vector NTI System Information is a utility application, which gathers data about the system on which Vector NTI is installed and consolidates that information into one file. Invitrogen Customer Support can use the System Information file to assist them in the diagnosis and resolution of system-related Vector NTI problems.

APIs

In Vector NTI, you can automate common tasks such as batch BLAST, batch annotation, batch primer design, etc. through Application Programming Interfaces. For more information, refer to the Invitrogen website, www.invitrogen.com/bioinformatics and the API User’s Manual.
Introduction

This chapter introduces you to the Vector NTI user interface and, in specific cases, to the Molecule Display Window, a powerful graphical tool for displaying and editing molecules. Included in this chapter are global operations used routinely in many of the Vector NTI Advance applications.

Vector NTI Advance 10 runs on a number of Windows platforms. All window manipulation techniques standard for given platforms can be used with the programs, such as moving and sizing windows, menus, toolbars, selecting objects and so forth.

This chapter assumes that you are familiar with Molecule Display windows and with basic display window techniques, such as maximizing windows, changing the sizes of text and gel panes, switching between panes, etc.

Launching Vector NTI

To launch Vector NTI, go to the Start menu on your computer. Select Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI. By default, the Vector NTI workspace is launched independently; however, the software allows you to configure the local Database Explorer to open simultaneously with the display window. From a Molecule Display window, select Edit > Options from the menu bar. On the General tab of the Options dialog box, select Open Local Explorer At Startup.

Online Help

The Online Help has been written to help you master the software.

In Vector NTI Advance applications, there are ways to open Online Help:

- Click the Help button ( ), then click on any toolbar button, opening the associated help topic.
• Click the Help button, then click on any command from the menu bar, opening the associated help topic.
• Press F1 from an open dialog box, opening its associated help topic.
• Select Help > Help Topics or Using Index from the menu bar of the application. In the Online Help that opens, you can browse through the Table of Contents or the Index, or launch a word search of the Online Help application.

If pressing F1 fails to open an Online Help topic, select Help > Help Topics, opening Online Help. Proceed with a Browse through the Table of Contents or Index or do a word search. Your topic may be in the Help files, but inadvertently not linked to its associated dialog box. Topics may be titled by their function rather than the dialog box name. For example, the New Molecule dialog box associated topic is named “Creating a New Molecule.”

The Application Workspace

A typical Vector NTI display window can include:

• Text information, including object descriptions, features, recombinant strategies and analysis results
• Graphical features of a molecule, a gel, or a contig assembly
• Analytical features of a molecule
• A pairwise or multiple alignment
• Appropriate sequences

From database or display windows:

• Local databases can be searched
• Public domain database searches can be launched through the Internet
• Text descriptions, graphical maps and sequences are edited
• Graphical representations may be formatted
• DNA/RNA and protein molecules and search results are stored in a database
• Preparation of publication-quality figures is supported.

Viewer Display Window Description

The display window title bar shows the name of the molecule on display or the name of the folder being displayed.

Every display window has a menu bar. Options in the submenus vary according to the specific application being used.

Shortcut menus associated with folders, molecules or objects from which they are opened provide many options for operations with those items. A shortcut menu is opened with a right click with the cursor paused over the related item.

All display windows contain two or more toolbars and variants relative to the active display pane.

• The Main Toolbar contains tools for major program functions, such as opening, saving, and printing Display windows, undoing and re-doing actions, and launching major subsections of the program such as the Database Explorer, etc.
• The Window Toolbar has several different variants according to the activated pane of the topmost display window.

Many menu and toolbar options are consistent throughout the Vector NTI software. Toolbar commands are often duplicated on the menu bar and shortcut menus, offering several means of accomplishing the same objective.
Note: Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

When you launch Vector NTI, the empty workspace is displayed. Upon opening a molecule in the Molecule display window, the following features can be viewed (see Figure 3.1)

**Viewer Panes**

Most display windows in Vector NTI Advance are divided into four sections, or panes:

- A Text Pane holds folders containing text descriptions and analyses results for a molecule shown in the Display window.
- A Graphics Pane displays a graphical depiction of a molecule. Features, restriction sites, motifs, etc. are shown by labeled symbols.
- An Analysis Pane allows certain types of DNA and protein sequence analysis properties to be performed and the results to be viewed as linear graphics.
- A Sequence Pane shows a nucleotide or amino acid sequence, and functional features of a molecule.

Variations of this will be explained in appropriate chapters.

![Figure 3.1 Molecule display window](image)

**Viewer Window Operations**

The following global operations for viewer windows apply in general to Vector NTI Advance applications. For details about operations specific to each application, refer to the appropriate chapters.

**Activating a Pane**

Click anywhere in a pane to activate it or click on the pane buttons appearing at the left of the lower toolbar.

**Note:** Clicking in a pane sometimes selects an entire molecule. To avoid this, use the toolbar buttons to activate the pane.
**Re-sizing Display Panes**
Split bars separate the Text, Graphics, and Sequence Panes. Re-size the panes by dragging the split bar that divides the panes. Position the cursor on the split bar until it changes to a two-headed arrow. LEFT-CLICK and drag to reposition the split bar.

**Scrolling the Text Pane**
Scroll bars appear when the contents of a pane exceed the viewing area. Move the scroll bars vertically or horizontally to reposition the pane contents for easier viewing. If the active pane is a Text Pane, it can also be scrolled with the arrow keys on the keyboard.

**Layout Options**
Select View > Maximize Pane to expand the size of an activated pane to full screen size. Select View > Restore Layout to return to the original pane layout.

Select View > Change Panes Layout to open the Pane Manager editor for modifying the arrangement of the panes in the workspace. For more details, see Pane Manager on page 22.

**Viewer Display Panes**

**Text Pane**
Text Pane folders contain descriptions of molecules and other objects, their characteristics and features, search results, methods and protocols.

Opening and Closing Folders
To open or close a folder, click the small square located to the left of the folder, double-click on the folder or line, select View > Toggle Folder, or select the folder and press ENTER.

Opening or closing a folder does not affect its state.

Expanding and Collapsing Folders
Opening of a folder simultaneously with some or all of its subfolders is called expanding that folder. Closing a folder simultaneously with some or all of its subfolders is called collapsing that folder. A folder and its subfolders are called a branch.

The View menu (and the shortcut menu) has commands for expanding and collapsing folders and subfolders in various combinations:

<table>
<thead>
<tr>
<th>Command</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expand One Level</td>
<td>Opens the focused folder; makes sure all subfolders are closed</td>
</tr>
<tr>
<td>Expand Branch</td>
<td>Opens the focused folder and all of its subfolders</td>
</tr>
<tr>
<td>Expand All</td>
<td>Opens every folder and subfolder in the text pane</td>
</tr>
<tr>
<td>Collapse Branch</td>
<td>Closes the focused folder and all of its subfolders</td>
</tr>
<tr>
<td>Collapse All</td>
<td>Closes every folder and subfolder in the text pane</td>
</tr>
<tr>
<td>Collapse Around</td>
<td>Closes all folders and subfolders, except any highlighted folders and their parents</td>
</tr>
</tbody>
</table>

*Table 3.1 View menu commands for expanding and collapsing folders and subfolders*

Corresponding toolbar buttons: **Expand Folder ( ), Expand Branch ( )** and **Collapse Branch ( )**
Linking the Display window panes

Linking the panes of a display window gives the Text Pane control over the information displayed in the Graphics and Sequence Panes. When linked, information from any open folder in the Text Pane is displayed in the Graphics and Sequence Panes, while information in closed folders is not displayed. Certain minimum information is always displayed, however; a molecule name and length in a Graphics Pane and the nucleotide or amino acid sequence in a Sequence Pane.

To link panes, activate the Text Pane and click the **Link Panes** button ( ) on the Window toolbar.

Linking panes reduces clutter, displaying only those objects that are important to your current task and lets you display symbols hidden by Vector NTI's Standard Arrangement of graphics objects. The Standard Arrangement of DNA/RNA molecules' graphics displays a maximum of 70 restriction sites plus 70 motifs, displayed to show as many different enzymes and motifs as possible.

You can override this display by linking the panes and opening the text folders for the only signals whose every occurrence you want to display.

Use the linked panes to create attractive publication-quality figures without switching to the Picture Editing mode and editing the picture manually. With the panes linked, close all folders, then open the folders of the signals you want to display, and then click the **Standard Arrangement** button ( ). Symbols and labels will be redistributed to take maximum advantage of available space.

Finding Objects on Graphics and Sequence Panes

Some objects, described in the Text Pane, are also displayed in the Graphics and Sequence Panes. To find the position of the object, select its line in the Text Pane and click the **Find** button ( ) in the window toolbar, select Edit > Find or press CTRL+ F. The object will be selected on the Sequence and Graphics Panes. The Find command is disabled if the selected line does not describe the object, displayed in two other panes.

**Note:** Vector NTI finds only single objects like a restriction site, not multiple objects like all restriction sites for a certain REN.

**Graphics Pane**

**Zooming the Graphical Map In and Out**

To enlarge a graphical map, activate the Graphics Pane and click the **Zoom In** button ( ) on the Window Toolbar or choose View > Zoom In. The keyboard shortcuts are ] and SHIFT + ]. To zoom in by a small increments, SHIFT + CLICK the **Zoom In** button.

To zoom out from a graphical map (reduce it by 50%), click the **Zoom Out** button ( ) on the Window toolbar or choose View > Zoom Out. The keyboard shortcuts are [ and SHIFT + ]. To zoom out by a small increments, hold down the SHIFT key and click the **Zoom Out** button.

If the current display is linear, change the horizontal zoom factor only using CTRL + CLICK on the **Zoom In** or **Zoom Out** button. Changing only the horizontal zoom factor changes the internal aspect ratio of the graphical display; this ratio is saved when you save Display Setup and can be applied to other Display windows as well.
To fit a graphical map to the current size of the Graphics Pane, click the **Fit to Window** button ( ![fit_window](image)) on the Window toolbar or choose the equivalent command in the View menu.

To zoom into a selection on the graphics map, making the selection the current size of the Graphics Pane, click the **Fit Selection to Window** button ( ![fit_selection](image)) on the toolbar or choose the equivalent command in the View menu. In some display windows, you can select **View > View Selection** or click the **View Molecule Fragment** button ( ![view_molecule](image)). To return to the original view, select **View > View Whole Molecule**.

One button is specific to the Gel Display window. Click the **True-Scale View** button ( ![true_scale](image)) to display the gel in an approximation of true scale. When you select this display option, Vector NTI looks up the pixel size (dot pitch) of your monitor and displays the gel based on that value.

**Sequence Pane**

### Formatting Sequence Text

To format sequence text, select it by click + dragging the mouse across the text. Because the Sequence Pane is active, you can use any of the following buttons or keystrokes to format the text.

<table>
<thead>
<tr>
<th>Button</th>
<th>Keystroke</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="B" /></td>
<td>CTRL + B</td>
<td>Set or clear bold character(s) attribute</td>
</tr>
<tr>
<td><img src="image" alt="I" /></td>
<td>CTRL + I</td>
<td>Set or clear italic character(s) attribute</td>
</tr>
<tr>
<td><img src="image" alt="U" /></td>
<td>CTRL + U</td>
<td>Set or clear underline character(s) attribute</td>
</tr>
<tr>
<td>Font style drop-down menu</td>
<td>--</td>
<td>Select font name for character(s)</td>
</tr>
<tr>
<td>Font size drop-down menu</td>
<td>--</td>
<td>Select font size for character(s)</td>
</tr>
<tr>
<td><img src="image" alt="font_color" /></td>
<td>--</td>
<td>Select font color for character(s)</td>
</tr>
<tr>
<td><img src="image" alt="background_color" /></td>
<td>--</td>
<td>Select background color for character(s)</td>
</tr>
</tbody>
</table>

*Table 3.2 Formatting sequence text*

### Miscellaneous Interface Features

#### Selection Techniques

**General Selection Techniques**

To select text in a Molecule Viewer display window’s Text Pane, click on the line you wish to select. To select multiple lines in a Sequence or Text Pane, press the mouse button and drag across the lines you wish to select.

To make a selection in a display window’s Graphics Pane, choose **View > Edit Picture** with an active Graphics Pane or click the **Edit Picture** button ( ![edit_picture](image)) to enter the Picture Editing mode and click on any object or label to select it. To select more than one object, click on each object in turn while holding down the **SHIFT** key.
To select text in a Sequence Pane, click and drag the mouse across the text.

You can select an entire pane by making that pane the active pane and choosing **Edit > Select All**.

**Broadcasting a Selection**

The Broadcast Selection feature of Vector NTI Advance allows you to select a region in the sequence pane of an opened molecule in one Vector Advance application and then transfer that selection to the same molecule that is currently open in other Vector NTI Advance applications.

The Broadcast Selection feature can be accessed from the shortcut menu visualized by right clicking on a selection in the sequence pane of any of the Vector NTI Advance applications (Figure 3.2):

**Figure 3.2 Broadcast Selection feature**

Choose the Broadcast Selection entry from the shortcut menu to mark that sequence area in the same molecule currently open in any other Vector NTI Advance application.

**Copying Display Window Text, Graphics or Sequences Image To The Clipboard**

Copying in Vector NTI can be done via several different methods:

**Edit > Camera or Camera Button**

The Camera feature copies the graphical representation, or “takes a picture” of the selection or activated pane contents. The contents are copied to the clipboard from where they are pasted into a new page or slide in other running applications.

Activate a pane or make a selection, then click the **Camera** button ( ) or choose **Edit > Camera**. In the dialog box that opens, choose the range and destination for the copied pane.

- **Range** – Specify the range of data to be copied. The selection option is unavailable if a selection is not defined in the active pane.
- **Copy To** – Specify the destination of the copy operation. If the File option is selected, the **Copy To** dialog box opens where the name of the output file and file type can be specified.

**Note:** If nothing is selected, the entire pane is copied. The format of this dialog box may vary slightly in different applications/panes.
Copying Sequence or Alignment Pane Data

In copying the contents of a Sequence or Alignment Pane to the clipboard, the Camera dialog box offers two formats for the copy:

- **Metafile picture** – only the snapshot of the Sequence or Alignment Pane will be copied to the Clipboard
- **Text** – you can further specify to print either the entire alignment or only the selected part of it (if there is selection at the moment). Sequences may or may not be wrapped. It can then be pasted into any text processor.

**Note:** A double-stranded sequence copied from Vector NTI becomes single-stranded when it is pasted.

The Clipboard stores Vector NTI text, sequence, alignment and graphical representations in both RTF (Rich Text Format preferred by word processors) and in plain text format (for e-mail and other applications preferring plain text).

**Note:** The RTF representation might not have exactly the same look as the sequence in the Sequence Pane due to limitations of RTF.

Vector NTI graphics copied with the Camera to a file can be stored as either a Windows Metafile (.wmf) or a Joint Photographic Experts Group (.jpg) file.

**Edit > Copy or Copy Button**

This copy feature transfers actual sequence and annotations, the molecule file, rather than just the graphical representation. Molecules copied this way can be shared and used in other Vector NTI Advance applications.

Text Pane data is copied as plain multiline text. Sequence data is copied as a plain text string. All sequence formats recognized by Vector NTI can be pasted from the Clipboard. Graphics Pane data is copied as a Windows Metafile.

Make the selection, then click the **Copy** button ( ) or select **Edit > Copy Sequence**. The data is copied to the Clipboard or to a file in text format. Preserves formatting and enables easy data exchange with other applications.

Copying Raw Sequence

To copy real sequence data to the Clipboard, select a sequence region and choose **Edit > Copy Sequence**. The sequence is copied in the plain text format without formatting or line numbering. This “raw” format is suitable for pasting into other molecular biology software including programs in Vector NTI Advance. The shortcut for the copy operation is CTRL + C.

**Edit > Copy To Command**

This copy command saves a whole molecule or selection of a molecule to the Clipboard or to a file in various formats, including saving a DNA/RNA sequence to a protein translation. If the whole sequence is chosen, you have a choice of saving to GenBank, EMBL, FASTA, plain sequence or amino acid translation (for DNA molecules). For a selection, the file format choices are FASTA, plain sequence or amino acid translation (for DNA selections).

Open a molecule, make a selection, if desired, and choose **Edit > Copy To > File** or **Edit > Copy To > Clipboard** from the pull down menu. Choose the desired format type and storage location on the hard drive, in the case of saving to a file. Choose Whole Molecule or Selection and Direct or Complementary strand (for DNA selections).

The **Edit > Copy To** command is implemented in Vector NTI and the Vector NTI Database Explorer.
Note: In the Vector NTI Database Explorer, you can select multiple molecules and copy them to the Clipboard in multiple GenBank, EMBL or FASTA format using the Copy To command.

Printing from Display Windows

Display windows are printed one pane at a time. To print a pane, activate it, then click the Print button ( ) on the main toolbar or choose File > Print.

Note: Print copies are similar to the screen display, so adjust the picture on screen before printing.

Page Setup

Text and Sequence Panes are printed in a standard format. For the text and sequence, you can change the content of what is printed, but not the layout. If a Text Pane is activated, any open folder is printed.

Graphical maps can be printed out at any size and in various positions on the page.

To change the size and position of a graphical map’s printout, select File > Page Setup. In the dialog box, set the margins (in inches) for your printout. Select File > Print Setup to specify the page orientation and size. Click OK to print.

Header and Footer Formatting Codes

The default settings for header and footer displays can be changed by selecting File > Page Setup. In the Page Setup dialog box, enter the appropriate formatting code in the header and footer sections. Click OK. (Default header settings are &D&c and default footer settings are &sPage&p&l.)

<table>
<thead>
<tr>
<th>Formatting Code</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>&amp;f</td>
<td>Prints the file name</td>
</tr>
<tr>
<td>&amp;D</td>
<td>Prints the document name (which is most cases is the molecule name)</td>
</tr>
<tr>
<td>&amp;p</td>
<td>Prints the page number</td>
</tr>
<tr>
<td>&amp;t</td>
<td>Prints the time representation</td>
</tr>
<tr>
<td>&amp;d</td>
<td>Prints the day of the month (as a decimal number)</td>
</tr>
<tr>
<td>&amp;s</td>
<td>Prints the date in full format</td>
</tr>
<tr>
<td>&amp;r</td>
<td>Aligns the text to the right side of the page</td>
</tr>
<tr>
<td>&amp;c</td>
<td>Aligns the text in the center of the page</td>
</tr>
<tr>
<td>&amp;l</td>
<td>Aligns the text to the left side of the page</td>
</tr>
<tr>
<td>&amp;u</td>
<td>Draws a line at the top of the page</td>
</tr>
<tr>
<td>&amp;L</td>
<td>Draws a line at the bottom of the page</td>
</tr>
<tr>
<td>&amp;b</td>
<td>Places a box around the text</td>
</tr>
</tbody>
</table>

Figure 3.3 Formatting Codes

Print Preview

To see a preview on your screen of what the printout will look like, choose File > Print Preview or click the Print Preview button ( ).

If you want to print just a selection, use the camera option. Select the region, click the camera on the Main Toolbar, paste the copied selection to a new application (such as word processing
software) and print from there. (See previous section on Copying Display Window Text, Graphics or Sequences Image to the Clipboard.)

Exporting Molecule Data

Vector NTI Advance applications allow simple export of a molecule as GenBank (GenPept for proteins), EMBL (SWISS-PROT for proteins), FASTA, or as text sequence. To export a molecule to an external file, choose File > Save As. In the dialog box, select the Save As File tab and designate the format type to be exported. Molecules and sequence information can also be exported by selecting File > Operations > Molecule Export. Select the Save As File tab and choose the file type you want to save and click OK in the standard File save dialog box.

Pane Manager

In most components of Vector NTI Advance, you can modify the layout of the display panes. To open the Panes Layout Editor, sometimes called Pane Manager, where you can modify display pane layout, select View > Change Panes Layout.

The main elements of the Pane Manager editor dialog box (Figure 3.4) are the hierarchy layout tree, the preview panel, and the drop-down menu (with a list of predefined layouts).

![Figure 3.4 Pane Manager editor dialog box](image)

The hierarchy layout tree represents the current layout (always the last settings used), while the node (Vertical, Horizontal, or Tabular) describes the relationship among the panes.

Options for modifying the pane layout appear on the shortcut menu, opened from the Vertical, Horizontal and Tabular nodes or from the Text, Graph or Sequence Pane leaves. Shortcut menu options are as follows:

- **Change Layout** – For Vertical, Horizontal, or Tabular nodes, this operation changes the relationship between underlying leaves (designated with green icons).
- **Dismiss** – This operation does not affect the root node. It removes the selected node, and shifts all of the underlying leaves to control of the root node. For example, if you dismissed Horizontal in the above graphic, the Text, Graph, and Sequence Panes would all be displayed vertically.
- **New sublayout** – This adds a new node in the designated area.
- **Up** – This will move the node or leaf up in the tree.
- **Down** – This will move the node or leaf down in the tree.

You can also change the parent node for any node or leaf. All you have to do is drag the node or leaf and drop it to new parent node.
The Predefined Layouts drop-down menu contains a list of predefined pane layouts.

The preview panel of the editor shows the layout for the current state of the hierarchy tree and is redrawn immediately when you change the layout. This panel is immediately redrawn after the state of the hierarchy tree is changed. The Info Pane is generally known as the Text Pane in most Vector NTI applications.

Click **OK** button after editing the layouts to apply your changes to the real panes or click **Cancel** to discard your changes.

**Notes:**
- You cannot add or delete a pane leaf. You can only move these leaves and configure their output.
- The Tabular selection places tabs at the bottom of each pane in the window.
- The Predefined option Standard returns the layout to the original.
Introduction

This tutorial chapter introduces you to the Vector NTI interface and Display windows, Vector NTI Advance’s primary means for displaying and manipulating DNA/RNA and protein molecules, gels, and other search and analysis features.

At the end of this tutorial session, you will be able to:

- Create and become familiar with Molecule Display windows for DNA and proteins
- Navigate the Vector NTI interface
- Manipulate molecule graphics, sequences, and Text Panes
- Review molecule data
- Select and examine molecule fragments
- Link molecule features for selective display in all panes
- Display amino acid translations of the nucleotide sequence and open corresponding proteins

Follow the steps of the tutorial in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch Vector NTI

To launch Vector NTI, select Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI.

The first time you enter Vector NTI after installation, the database is empty. Vector NTI asks permission to import an initial set of database objects from Vector NTI archives. Click the OK button in the confirmation dialog boxes for the creation of all objects.

The DNA molecules, proteins, enzymes, oligos, and gel markers are imported to form the initial Vector NTI database.
2. Inspect the Vector NTI Workspace and Database Explorer Windows

Vector NTI first appears on the screen with an empty workspace.

![Figure 4.1 Vector NTI Empty Workspace](image)

To open a molecule, access the Database Explorer from the toolbar icon ( ) or the Molecule menu and select a DNA/RNA molecule or Protein from the database table. The Vector NTI database is a collection of molecules and laboratory data, organized for easy retrieval and management in Vector NTI’s Database Explorer.

The Explorer window can be closed without exiting Vector NTI. You can activate or reopen the Database Explorer window at any time using the **Local Database** button ( ) on the main toolbar of the Vector NTI workspace.

You will work extensively with the Database Explorer tutorial, Chapter 7, but for now close the Explorer by clicking on the **Close** button ( ).

**Note:** By default, the Vector NTI workspace opens independently, however, it can be configured to open the local Database Explorer simultaneously each time a Vector NTI session is initiated. From a Molecule Display window, select **Edit > Options** from the menu bar. On the **General** tab of the Options dialog box, select **Open Local Explorer At Startup**.

3. Create and Inspect a Molecule Display Window for pBR322

A Molecule Display window displays text information including a description of a molecule, features and a graphical map of the molecule and its sequence. Before proceeding any further however, let’s open a molecule so that you can review the features as they are discussed.

On the Molecule Display main (upper) toolbar, click the **Open Document** button ( ). In the Open dialog box, select the Database DNA/RNAs tab. The **DNA/RNA Molecules (MAIN) data-**
base is selected in the drop-down menu. On the molecule list, scroll to pBR322 and double click on it, opening its Molecule Display window.

Figure 4.2 Molecule Display window for pBR322

A Molecule Display window title bar displays the name of the open object.

The Molecule Display window is divided into three panes: a Text Pane (upper left) containing folders with descriptions of a molecule and analysis results, a Graphics Pane (upper right) displaying a graphical depiction of the molecule map including labeled features and restriction sites, and a Sequence Pane (lower) displaying the molecule’s nucleotide or amino acid sequence.

All Vector NTI Advance windows have a Main Menu and toolbars where many different options and operations can be initiated. While many toolbar buttons are consistent in all display windows, many toolbar buttons are context-sensitive. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

4. Arrange the Display Window Conveniently

To modify the relative sizes of the display panes, you can move the vertical and horizontal split bars. Move the cursor arrow to the split bar where it changes to a two-ended arrow. CLICK + DRAG with the left mouse button to move the split bar. Additionally, each pane has its own scroll bars when the pane content exceeds its display area.

To activate the different panes of the display window, click the Text Pane ( ), Graphics Pane ( ) or Sequence Pane ( ) buttons on the Window Toolbar. (You can also activate each pane by clicking it.) Note how the Windows Toolbar changes, giving you different tools for each pane.

Switch to the Graphics Pane. Move the horizontal and vertical split bars to enlarge the Graphics Pane viewing area.
Click on each of the following buttons to resize the graphical map:

- **The Zoom In** button ( ) enlarges the image
- **The Zoom Out** button ( ) reduces the image
- **The Fit to Window** button ( ) fits the image to the size of the window.

Now reduce or increase the map size in increments by holding down the SHIFT key and clicking the **Zoom** buttons. Using the scroll button on your mouse while holding down the SHIFT key has the same effect as clicking the **Zoom** buttons.

5. **Select a Fragment or a Feature in the Graphics Pane**

Now let’s study Vector NTI’s selection techniques.

Make sure the Graphics Pane is active. Choose **Edit > Set Selection** on the menu bar, opening the Set Selection dialog box. Enter the range 100 bp – 1000 bp and click **OK**.

A selection wire-frame highlights the fragment on the molecule map (Figure 4.3), with a caret (short darkened line perpendicular to the wire-frame) marking one end of the wire-frame.

The selection position is displayed in the status bar at the bottom of your screen. The caret position is also indicated on the status bar. The selection is simultaneously highlighted in the Sequence Pane.

![Figure 4.3 A selection wire-frame highlights the fragment on the molecule map](image)

### Modifying the Selection

Click on the status bar with the mouse, opening the Set Selection dialog box again. Change the setting to 200 bp – 2000 bp.

You can reposition the 5’ or 3’ ends with the mouse by “dragging and dropping”. Move the cursor to the 5’ end, identified when the cursor changes shape to a crosshair with the 5’ mark ( ). Click and drag the 5’ caret to a new position near the start of the molecule (“12:00 o’clock”).

Hold the **SHIFT** key down and press the right or left keyboard arrows. This moves the cursor 1 nucleotide (NT) at a time. **Follow this in the Status Bar.** Move the 5’ end to “1 bp”.

---

**Vector NTI Advance 10 User’s Manual**

28
Press the END key to move the caret to the 3’ end (or click the cursor on the 3’ end). Press and hold the SHIFT key while moving the left or right arrows to modify the selection. Move the 3’ end to 1250 bp.

Hold down the SHIFT + CTRL keys and press the right or left arrow several times. Note that the caret now moves 10 NT at a time. Release the CTRL key, hold down SHIFT and press the left arrow. The caret is moving 1 NT at a time again. If you hold down the arrow keys while holding the SHIFT key, the caret moves rapidly through the sequence. Release the SHIFT key and press HOME. The caret returns to the 5’ end.

To make a new selection, click anywhere in the blank area inside the molecule map, selecting the entire map. Click on the caret and drag the wire-frame clockwise. Release the mouse.

Tip: If you wish to deselect everything and are having problems doing so in the Graphics Pane, click anywhere in the Sequence Pane.

Selecting Features

Locate the tetracycline resistance gene, TC(R), on the graphics map. Pause the cursor over the functional segment or label for a moment. A pop-up message with details about the functional signals appears under the cursor.

To simplify selection of a functional signal on a graphical map, the paused cursor changes to a hand ( ). Click once over the TC(R), selecting the gene. A wire-frame selecting the gene appears and the selected signal changes color. The gene sequence is concurrently selected in the Sequence Pane.

Selection Techniques Summary for the Graphics Pane

When making or altering fragment selections, there are several general points to remember:

- Click and drag the mouse to select a region.
- Click on a feature to select it.
- Press SHIFT + right or left arrow to adjust an existing selection one nucleotide at a time.
- Press CTRL + SHIFT + right or left arrows to move the caret 10 nucleotides at a time.

Note: A selected fragment is simultaneously highlighted on the Sequence Pane. Vector NTI always displays selection in both panes no matter where you make the selection.

6. Examine pBR322’s Nucleotide Sequence

Now let’s work with pBR322’s nucleotide sequence.

Drag the horizontal split bar up for better viewing of the nucleotide sequence. The nucleotides are displayed in 10 NT blocks. Restriction sites appear by name above their recognition sites. Motifs and ORFs also can be displayed by turning those options on the Molecule Display Setup, to be discussed later in this chapter.

Click anywhere in the Sequence Pane to activate it. Any prior selection disappears and the blinking caret is placed at the position where you placed the cursor.
When the DNA/RNA Display Window opens, the displayed sequence is double stranded. Click on the Show Both Strands (Double/Single Strand) button to change to a single-stranded sequence display (Figure 4.4).

Scroll to the top of the sequence, using the scroll bar to the right of the sequence. Select the first few lines of the sequence by clicking and dragging the mouse left to right along the line and down through the sequence. The nucleotides are highlighted as you drag and their coordinates are displayed in the selection box on the status bar. The selection also shows within a wire-frame in the Graphics Pane. Release the mouse.

To translate the selected sequence into amino acids, click the Translate Direct and Translate Complementary buttons. The corresponding amino acids appear above (direct strand) and below (complementary strand) the highlighted sequence (Figure 4.5).

Numerals to the left of the sequence row indicate the phase of the translations. For the Direct strand translation, phases are determined by the position of the first selected nucleotide relative to the beginning of the molecule and are designated +1, +2, or +3. Complementary strand translation phases are determined by the position of the last selected nucleotide relative to the end of the molecule and are designated -1, -2, or -3.
To change from the 3-letter amino acid code to a 1-letter code, choose the Display Setup button \( \text{Display Setup} \) and select Display Setup on the drop-down menu.

![Figure 4.6 Molecule Display Setup dialog box](image)

In the Molecule Display Setup dialog box (Figure 4.6), choose the Sequence Setup button and then the 1-letter code radio button in the Sequence Setup dialog box. Click OK in both dialog boxes. Inspect your selection. This dialog box is also used to modify or specify the display of other features such as restriction sites, motifs and ORFs.

Click on the Erase Translations button \( \text{Erase Translations} \) on the toolbar to erase the translations.

7. Work with pBR322’s Text Description

Click in the Text Pane or click the Text Pane button to activate the Text Pane. Drag the vertical split bar to the right to enlarge the Text Pane.

The Text Pane is made up of folders that can be opened or closed by double clicking or clicking on the \( \text{+} \) to the left of the folder name. Open and review the contents of each of the folders.

Note particularly the following:

**General Description Folder:** This folder contains the molecule description. Close the folder by clicking on it.

**Feature Map Folder:** Select this folder and click the Expand Branch button \( \text{Expand Branch} \), opening all of the Feature map subfolders. Move through the subfolders until you come to the TC(R) in a subfolder of the CDS (Coding DNA Sequence) folder. Note the coordinates shown.

Select the TC(R) folder and click the Find button \( \text{Find} \). This selects the gene in the Sequence and Graphics Panes. Close the folders with a double click.

The Find button can also be used to locate ORFs, motifs and other objects. Select the appropriate folder in the Text Pane and press the Find button again. The keyboard equivalent is CTRL + F.

**Restriction Map Folder:** Click on this folder to open it. Click on the Expand Branch button \( \text{Expand Branch} \) in the Window Toolbar, opening all restriction map subfolders.

Each subfolder contains a restriction site, listed alphabetically. Restriction sites that are present at least once on pBR322 have active (colored) folders showing all of the site’s positions on the
molecule and its recognition sequence. To the right of each enzyme’s name, its recognition sequence and cleavage points are displayed. Inactive folders (grayed out) are for restriction endonucleases with no sites on the molecule.

The set of restriction enzymes listed in the Text Pane and searched for on all DNA molecules when you open them is defined in the Molecule Display Setup dialog box. To examine the list of selected enzymes, click the **Display Setup** button ( ) on the Window Toolbar and choose **Display Setup** on the drop-down menu. Click the **RMap Setup** button, opening the Restriction Map Setup dialog box (Figure 4.7).

![Figure 4.7 Restriction Map Setup dialog box](image)

In this dialog box, you can add (or remove) any enzymes you wish to be searched for on all open DNA molecules. Do not change the default settings. Click **Cancel** in both dialog boxes to return to the Display window.

Locate the **ApaLI** folder in the Text Pane. The folder is open, showing a list of all **ApaLI** sites on the pBR322 molecule. Click on the first site line to select it and click the **Find** button ( ) on the Window Toolbar. The **ApaLI site is selected on the Graphics Map and in the Sequence Pane**. Move to those panes to verify this, then return to the Text Pane. Close the Restriction Map Folder with a double-click.

### 8. Link pBR322’s Text Pane to the Graphics and Sequence Panes

In Vector NTI’s Display Windows, you can link panes so that the Graphics and Sequence Panes display only those objects whose folders are open in the Text Pane.

Click the **Link Panes** button ( ) on the Window Toolbar. Most of the information disappears from the graphical map, leaving only the molecule name and length. In the Text Pane, open the Feature map folder and only the subfolders for the TC(R) gene (in the CDS subfolder) and the Restriction Map folder with only the following enzyme folders open: **ApaLI**, **AvaI**, **BamHI**, **EcoRI** and **PstI**. Because only open folders are now linked with the Graphics Pane, only those features are illustrated in the Graphics Pane. (Make sure the folders are closed for any features you do not want displayed.) Note that the sites and functional signals appear on the graphic map as you open them in the Text Pane. The sequence fragments are also highlighted at the same time.
Activate the Graphics Pane and click the **Standard Arrangement** button on the Window Toolbar ( ) to rearrange the restriction site labels for optimum display (Figure 4.8).

[Figure 4.8 Standard arrangement of restriction site labels for optimum display]

Turn off the text/graphics link by clicking the **Link Panes** button in the Window toolbar. Now the graphics and Sequence Panes are complete and unaffected by opening or closing text folders.

9. **Print pBR322's Text Description, Graphical Map, and Sequence**

To print the Text Pane contents, first make sure the Text Pane is active. Choose **View > Expand All** from the menu bar. This opens all folder and subfolders.

Click on the **Print Active Pane** button ( ). The contents of all open folders in pBR322's Text Pane are printed.

Print the graphics and Sequence Panes in the same manner, activating each pane first. The *pBR322's sequence printout consists of five pages of a standard format.*

10. **Create a Display Window for 41BB_HUMAN**

Click on the **Local Database** button ( ) on the Main Toolbar to open Database Explorer.

In the Database Table drop-down menu in the upper left, click on the arrow and select the Protein Molecules table. Select the Protein Molecules (MAIN) subset in the All subsets window.
Double-click on the 41BB_HUMAN molecule in the Database Protein Molecules list in the right pane. A Molecule Display window opens for 41BB_HUMAN (Figure 4.9) containing its text description, analysis results, graphical Feature map, and sequence:

![Figure 4.9 Molecule Display window for 41BB_HUMAN](image)

Double-click on the Analysis folder in the Text Pane, opening two tables showing the results of the automatic protein analysis.

Select both tables by opening a shortcut menu with a right click on the Analysis Folder. Choose the **Select Folder** command.

Click the Camera button or choose Edit > Camera on the menu bar. In the Camera dialog box, choose the range Selection and destination of your data Clipboard.

Click the Copy button, switch to a word processor program and paste the result into a new or existing document. *The results are displayed in a standard tabular form as shown here, (in a truncated version):*

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Entire Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>255 aa</td>
</tr>
<tr>
<td>Molecular Weight</td>
<td>27897.66 m.w.</td>
</tr>
<tr>
<td>1 microgram =</td>
<td>35.845 pMoles</td>
</tr>
<tr>
<td>Molar Extinction coefficient</td>
<td>11250</td>
</tr>
<tr>
<td>1 A[280] corr. to</td>
<td>2.48 mg/ml</td>
</tr>
<tr>
<td>A[280] of 1 mg/ml</td>
<td>0.40 AU</td>
</tr>
<tr>
<td>Isoelectric Point</td>
<td>8.13</td>
</tr>
<tr>
<td>Charge at pH 7</td>
<td>3.72</td>
</tr>
</tbody>
</table>

*Table 4.1 Results of an automatic protein analysis in a word processor program*
11. Create a Display Window for 1B14_HUMAN

Return to the Database Explorer window, the Protein Molecules table and select the Protein Molecules (MAIN) subset.

Double-click the 1B14_HUMAN protein in the molecules list, opening it in the Molecule Display window. Arrange the panes conveniently.

Note that with a large Feature map, the Graphics Pane appears crowded. To make it more convenient to work with, use the “link mode” described above in Step 8 to display fewer features. Alternatively, Vector NTI allows you to limit the view by choosing a fragment of the molecule for closer inspection (Figure 4.10).

Select the region 231 aa – 276 aa using the techniques described for DNA molecules.

Click on the View Molecule Fragment button ( ) and choose View Selection (231 aa – 276 aa) from the popup menu. All three views are rearranged to display only the selected fragment.

<table>
<thead>
<tr>
<th>Amino Acid(s)</th>
<th>Number Count</th>
<th>% by weight</th>
<th>% by frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Charged (RKHYCDE)</td>
<td>83</td>
<td>36.68</td>
<td>32.55</td>
</tr>
<tr>
<td>Acidic (DE)</td>
<td>25</td>
<td>10.85</td>
<td>9.80</td>
</tr>
<tr>
<td>Basic (KR)</td>
<td>29</td>
<td>14.43</td>
<td>11.37</td>
</tr>
<tr>
<td>Polar (NCQSTY)</td>
<td>90</td>
<td>34.08</td>
<td>35.29</td>
</tr>
<tr>
<td>Hydrophobic (AILFWV)</td>
<td>67</td>
<td>27.06</td>
<td>26.27</td>
</tr>
<tr>
<td>A Ala</td>
<td>11</td>
<td>3.02</td>
<td>4.31</td>
</tr>
<tr>
<td>C Cys</td>
<td>25</td>
<td>9.33</td>
<td>9.80</td>
</tr>
<tr>
<td>D Asp</td>
<td>11</td>
<td>4.51</td>
<td>4.31</td>
</tr>
<tr>
<td>E Glu</td>
<td>14</td>
<td>6.34</td>
<td>5.49</td>
</tr>
<tr>
<td>F Phe</td>
<td>16</td>
<td>8.14</td>
<td>6.27</td>
</tr>
</tbody>
</table>

Table 4.2 Results of an automatic protein analysis in a word processor program
Double-click on the Analysis folder in the Text Pane. Note that the analysis tables now show values for the selected fragment as well as for the entire protein.

Most of the other techniques for manipulating Protein Molecule Display windows are exactly the same as for DNA/RNA Display windows.

12. Close the Display Windows and Exit Vector NTI

To complete your first session with Vector NTI, select File > Close on the menu bar, closing the display windows.

To exit Vector NTI, select File > Exit.
Introduction

This chapter introduces you to operations of Molecule Display Windows that allow you to edit a molecule’s Feature map and sequence. Although in this chapter you will work only with DNA molecules, all operations are applicable to protein molecules as well.

At the end of this tutorial session, you will be able to:

- Create a Molecule Display window for a molecule
- Edit the general data, Feature map, and sequence for a DNA molecule
- Rename, describe and save a molecule in the database
- Alter starting coordinates for a DNA molecule

1. Launch Vector NTI

Launch Vector NTI by selecting Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI.

2. Open a Display Window for pBR322

Open a Molecule Display window for pBR322 by selecting File > Open from the Main menu or click on the Open Document ( ) button. Select the Database DNA/RNAs tab in the Open dialog box and scroll down to select pBR322 by double-clicking on the molecule name.

Alternatively, access the Database Explorer by clicking the Local Database ( ) button and choose the DNA/RNA molecules table from the pull-down menu. Select the DNA/RNA Molecules (MAIN) subset and double-click on the pBR322 molecule.

A display window opens in the workspace, containing pBR322’s text description, graphical features, and restriction maps, and sequence.
3. Edit pBR322's General Data

At the top of the Text Pane, double-click the pBR322 name. This opens the Edit pBR322 dialog box (Figure 5.1):

![Figure 5.1 Edit pBR322 dialog box (DNA/RNA Molecule tab shown)](image)

First, let’s add a keyword to the list of pBR322’s keywords. Choose the Keywords tab and type in *My own plasmid* in the New Keyword text entry box. Click the < Add button. Now the keyword list contains two keywords: “ATCC” and “My own plasmid”. Switch to the DNA/RNA Molecule tab, and change the Description of the molecule to “My pBR322.” Click the OK button.

Vector NTI updates the General Description folder in the Text Pane to show the new properties. An asterisk now appears on the title bar, indicating that the molecule in a window is a modified version of the pBR322 molecule stored in the Vector NTI database.

To save the modified pBR322 into the database under a new name, select File > Save As from the menu bar and enter *My pBR322* to the DNA/RNA text box. Click OK.

The modified molecule is stored under the new name and updated in the display window. The title bar and both the text and Graphics Pane show the new name “My pBR322”. The title bar is no longer marked as modified (asterisk) because the displayed molecule is exactly the same as the database molecule “My pBR322”.

Open a Display window for pBR322 to make sure that the original molecule was not modified.

To do so, click the Open Document button ( ) on the Main Toolbar and select pBR322 from the Database DNAs/RNAs tab. To continue, close all extra Display windows except “My pBR322”.


4. Edit My pBR322’s Sequence

Activate the Sequence Pane and select the fragment 21 bp–40 bp (Figure 5.2) using either the drag and select method or the Set Selection dialog box described in the previous chapter. Note that the selected sequence contains CiaI and HindIII sites.

Choose **Edit > New > Replace Sequence 21 bp–40 bp** from the menu bar. In the Replace Sequence dialog box (Figure 5.3), use arrow keys to move the caret to the position "before 26 bp" (the caret position is displayed at the bottom of the dialog box). Press the **Backspace** key twice to delete C and T to the left. Type **A** twice so that the resulting sequence is TCAAA-GATAAGCTTTAATGC. The status line on the bottom of the dialog box displays "inserted 2, deleted 2" message.
Click the **OK** button to finish editing the fragment. In the updated Display window (Figure 5.4), the Clai site on the modified fragment has disappeared.

![Figure 5.4 Display window with updated sequence](image)

### 5. Revert to Database Version of My pBR322

The “My pBR322” molecule in the display window is now flagged (with an asterisk) as modified because it differs from the version of “My pBR322” currently saved in the database.

To return to the database version, choose **File > Revert To Saved** from menu bar. Click **OK** in the confirmation dialog box.

The molecule in the display window returns to its database state and the window is no longer marked as modified (no asterisk). Notice that Clai site returned to its place in the original sequence.

### 6. Insert a New Sequence Fragment

In general, editing a molecule sequence requires the adjustment of its Feature map. In this step, we will insert a new sequence fragment into our molecule and look at the resulting Feature map.

Before you start, notice that the current AP(R) position is 3293 bp–4156 bp; the TC(R) position is 86 bp–1276 bp. To see the positions of the features, pause the mouse cursor over the feature or its label on the graphics map. A popup label appears, displaying the name and position of the feature under the cursor.

Place the caret at 200 bp using the mouse, arrow keys, or **Edit > Set Caret Position**. Choose **Edit > New > Insert Sequence at 200 bp** from the menu bar, opening the Insert Sequence dialog box. Type in 10 Ts (the number of nucleotides you entered is shown at the bottom of the dialog box). Click **OK** to insert the sequence.

Vector NTI figures out that the modification you made requires an adjustment in the Feature map. The dialog box appears with the message “CDS TC(R) is affected by sequence editing.” Click the **Keep** button to leave the feature in a Feature map. Note that the inserted nucleotides are highlighted in the sequence (Figure 5.5).
Place the mouse cursor over AP(R) again to determine its position. Note that AP(R) has been moved 10 nucleotides clockwise and is now positioned at 3303 bp–4166 bp.

Vector NTI automatically moves all features located after the insertion but it does not move any features before the insertion. If the insertion point is inside a feature and you decide to keep the feature, Vector NTI automatically moves only the 3’ end of the feature. Observe that the 3’end position of TC(R) was changed to 1286 bp.

7. Edit the TC(R) Signal

Double-click on the TC(R) signal or its name in the Graphics Pane. (The mouse cursor changes to a hand over features and their names). Alternatively, click on the signal once and choose **Edit > Feature Properties** from the menu bar or choose **Feature Properties** from TC(R)’s shortcut menu. *(To display a shortcut menu for any object, right click on the object. The shortcut menu contains commands appropriate for the item to which you are pointing.)*

In the Molecule Feature dialog box that opens, you can change the name, type or description of the feature as well as its position on the molecule. Change the name of the signal to “Old TC(R)” and its description to “10 bp fragment inserted”. Click the **OK** button. Vector NTI updates the Display window to show the new name of the signal (Figure 5.6).
Vector NTI allows you to easily add and delete features in a Feature map. To delete the P2_P promoter from the Feature map, select it with a click on the label in the Graphics Pane and choose Edit > Delete Feature From FMap from the menu bar. Alternatively, select P2_P and choose Delete Feature From FMap from the shortcut menu. Click the OK button in the confirmation dialog. P2_P is deleted from the Feature map and disappears from the Display window.

Let's add a new feature to the Feature map of My pBR322. First, select region 3000bp–3500 bp and click the Add Feature button ( ) on the Window toolbar. Alternatively, you can choose Edit > New > Add Feature to FMap. In the Molecule Feature dialog box, Vector NTI puts the currently selected region into the From and To fields. By default, the “Misc. Feature” type is assigned to new DNA/RNA features in the Feature Type panel, but you may change the type to any other type. Name the new feature New Feature and click the OK button. The new feature appears in the text and Graphics Panes of the Display window (Figure 5.7).

To save the modified My pBR322 into the database, choose File > Save As from the menu bar. Click the OK button, then OK again to overwrite the version of My pBR322 currently stored in the database.

9. Change the Starting Coordinate of My pBR322

The Molecule menu options control the operations that can be performed on a molecule as a whole. Let's change the starting coordinate of My pBR322 so that all features after the inserted fragment have the same coordinates they have in the original pBR322 molecule. Choose File > Molecule Operations > Advanced (DNA/RNA) > Change Starting Coordinate from the menu bar. In the dialog box that opens, enter a new starting coordinate. Since the length of the inserted fragment was 10 bp, the new starting coordinate should be 1 bp + 10 bp = 11 bp. Enter 11 into the New Start field and click the OK button. Click OK in the confirmation dialog. Vector NTI recalculates the Feature map and updates the Display window. Verify that the new coordinates of AP(R) feature are the same as in the pBR322 (Figure 5.8): 3293 bp–4156 bp and “Old TC(R)” is now at 76bp–1276bp.

Notice that the Display window is not marked as modified. For all commands from the Molecule menu, Vector NTI always operates on the molecule in the database and updates the Display window if required. If a molecule in a Display window was modified but not saved at the time you
chose a command, Vector NTI asks you to save the modified molecule into the database and then repeat the command.

Figure 5.8 New coordinates of AP(R) feature

10. Close the Display Window and Exit Vector NTI

You have finished your session with Vector NTI. Close the Molecule Display window using the Close command in the File menu. Quit Vector NTI with the File menu’s Exit command.
Introduction

In this chapter, you will learn to edit the representation of molecule graphics maps in a Molecule Display window. Although here you work only with DNA molecules, all techniques described in this chapter work equally well for protein molecules.

There are four ways to modify the text and graphics in the Graphics Pane. At the end of this tutorial session, you will be able to:

- Use the Arrangement Setup Palette to modify the standard display of a molecule
- Modify the Graphical Display Setup for a DNA molecule and save the settings for use with other molecule displays
- Format object graphics and labels manually in Picture Editing mode
- Define label and feature styles for a graphical image
- Save the molecule’s graphical representation to a Molecule Document file.

Note: The operations in this tutorial have no effect on the molecule itself, but only on its graphical display.

1. Launch Vector NTI

Launch Vector NTI by selecting Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI.

2. Open pBR322 in a New Display Window

In the active workspace, click the Open Document button ( ) on the Main Toolbar. In the Open dialog box, select the Database DNAs/RNAs tab and the DNA/RNA Molecules (MAIN) subset. Scroll to pBR322 and double click on it to open a new Molecule Display window.
3. Arrange the Display Window Conveniently

Click the **Graphics Pane** button ( ) to activate it. Use the split bars and scroll bars to get a convenient display of pBR322’s graphics map. Now **SHIFT + CLICK** the **Zoom In** ( ) and **Zoom Out** ( ) buttons on the window toolbar to resize the graphics map until its labels are easily readable.

4. Change the Arrangement Setup

Vector NTI automatically rearranges the symbols and labels on the graphics map when you change the set of restriction enzymes, edit the functional map of the molecule, etc. This is called the Standard Arrangement.

To change the parameters of Standard Arrangement, press **CTRL** while you press the **Standard Arrangement** button ( ) on the Window Toolbar. The Arrangement Setup Palette appears, where you can change the standard arrangement settings like size of labels and width of signals.

Experiment with the Standard Arrangement formatting by pressing several of these buttons. The arrow buttons alter the molecule features display, while the “A” buttons modify the text display. When you are satisfied with the arrangement results, click the **Standard Arrangement** button again to close the palette.

5. Change the Graphics Setup of CDS Signals

Using the Graphics Display Setup dialog box, you can change the predefined display styles for elements of feature and restriction maps. You can specify that newly defined styles be available for all features or that they be dedicated to only a specific feature type.

To modify the way CDS (Coding DNA Sequence) signals are displayed, right click on one of the CDS signals in the Graphics Pane and choose **CDS Display Setup** from the shortcut menu. In the Graphics Display Setup dialog box, the CDS feature type is automatically selected. Here you will specify that all features of the same type (CDS) are formatted in the same way.

---

**Figure 6.1 CDS Display Setup**
Using the check boxes, you can even choose whether or not to display labels for a selected feature type. You can choose label styles including font color and text format, and symbol shapes and styles.

Click the More button near the Label Style drop-down menu. In the Properties dialog box that opens, click the Font button. In the Font dialog box, choose Bold font style, and scroll to the blue color and select it. Press OK. Now choose the Line tab on the Properties dialog box and select a shade of blue to match the font color. Click OK.

If this is the first time you have created a style profile, you are asked whether you want to create a shared style to be used for all feature types or a specific style that would be used for CDS signals only. Press No to create a unique style named “CDS Label”. This means that this style is available to format only CDS features. This is indicated now in the Label Style box by the term CDS Label (Figure 6.2).

To change the symbol style, click the More button near the Symbol Style box and choose the Fill tab. On the Fill tab, select a light blue color. Click OK to close the dialog box. When asked what kind of style you want to create, choose Yes to create a shared style. Vector NTI assigns a sequential number to the new style (Shape 5) and allows you to use this style to format not only CDS signals but other features as well.

Press OK. This closes the Graphics Display Setup dialog box and automatically rearranges pBR322’s Graphics Pane to show the new display style for CDS signals (Figure 6.3).

The style modification you just made has no effect on the molecules displayed in other Display windows. Vector NTI associates a style sheet with every Display window in the system. A style sheet is a collection of different kinds of display settings such as Graphics display settings (like those you just modified), sequence settings like 1-letter or 3-letter translations and Feature map settings.

All settings in the current style sheet are accessible from the Molecule Display Setup dialog box you examined in the previous tutorial.

To save your modified style sheet, making it easily available to use with other molecules, click on the Display Setup button ( ) on the Window toolbar and choose the “Save Settings As” command from the popup menu. Type the name Blue in the Save As dialog box and click OK. To save unused styles, click No.
Close all display windows and reopen the pBR322 window. When first displayed, it uses the Default style sheet. Click on the Display Setup button on the Window toolbar. The list of available style sheets in the drop-down menu now has both elements: “Default” and “Blue”. Choose the “Blue” style sheet, applying the style you just defined to the Display window. CDS signals are now shown in blue.

6. Turn on the Picture Editing Mode

If you prefer, you can format graphical maps manually and change the display style for individual features and elements of other graphical maps.

The Vector NTI Graphics Pane can operate in two different editing modes:

- **Molecule Editing** mode (the default) where you can modify a molecule itself, such as adding functional features, restriction sites or inserting or deleting residues in a molecule. *This was covered in Chapter 5.*
- **Picture Editing** mode where you can format individual labels, arrows and symbols of a molecule’s graphical maps. Operations on the graphics in this mode do not modify the molecule itself.

To switch the Graphics Pane into the Picture Editing mode, activate the Graphics Pane and click the Edit Picture button ( ) on the Window Toolbar.

**Note:** Vector NTI automatically reverts out of Picture Editing mode back to Molecule Editing mode as soon as you select a sequence fragment in a Sequence Pane or perform some operation on a molecule.

7. Turn the TC(R) Arrow to Blue Cross-Hatching

Locate the arrow representing the TC(R) (tetracycline resistance gene) symbol on the graphics map. Let’s change its current pattern to cross-hatching and its color to light blue, and then enlarge the arrow.

Verify that the Picture Editing mode is on. (View > Edit Picture should display an enabled Picture Editing Mode button.) Point the cursor at TC(R) (the cursor changes shape to a pointing hand) and double-click on the arrow. In the Properties dialog box, click on the Fill tab. On the Fill tab, click in the pattern box. Select the crosshatched pattern in the scroll-down list, and click on it (Figure 6.4).

![Figure 6.4 Selecting the cross-hatched pattern in the Pattern field of the Properties dialog box](image)

In the color selection box, choose light blue and click on OK. Immediately after the color change, the arrow’s colors are inverted, showing that the signal is currently selected. Click in the graphical map outside the TC(R) arrow, deselecting it. The new pattern and color are displayed.
8. Enlarge the TC(R) Arrow

To enlarge the TC(R) arrow, click on it, selecting it. "Drag boxes" or "handles" now appear at its corners. **Note:** If you do not see drag boxes after you click on the arrow, the pane is not in Picture Editing mode.

![Figure 6.5 Making the arrow thicker and thinner by dragging the handles](image)

Move the cursor to the drag boxes at the tail of the arrow where the cursor turns to a thin four-pointed arrow ( krat ). Click and hold the left button on one of the drag boxes, and move the mouse a bit. Practice making the arrow thicker and thinner by dragging the handles (Figure 6.5). **The drag boxes at the front of the arrow are for changing the shape of the arrowhead. Feel free to experiment with them on your own.**

To drag the arrow radially inwards or outwards on the map, move the cursor to the middle of the TC(R) arrow where it changes to a thick white four-pointed arrow ( krak ). Click and hold the left button and drag mouse inward or outward. **CTRL + SHIFT + left button-drag moves the arrow around the circle.**

Remember, although the functional symbol graphics may have changed position on the graphics map, the actual molecule remains unaffected because you are in Picture Editing Mode. The double key combination has been designed with the drag to prevent you from moving functional signals out of position on the graphical map by accident.

To undo any changes you make in the graphical map, click the **Undo** button on the main toolbar ( ![undo](image)), or press **CTRL + Z**.
9. Format TC(R)'s Label

To change the formatting of the TC(R) label, double-click on it, opening the Properties dialog box for the label (Figure 6.6).

Select the Text tab and press the Font button. Set the font to 18-point bold italic in the Font dialog box (Figure 6.7). Click the OK buttons in both dialog boxes. Reposition the label using LEFT-CLICK + drag. Practice selecting and dragging other labels around the screen.

Undo the reposition changes by clicking the Standard Arrangement button ( ) on the toolbar. If the picture was modified, Vector NTI displays a confirmation dialog before proceeding.

10. Add a Text Annotation

To add a text note indicating that TC(R) is to be cloned into pUC19, click the Add Annotation button ( ) on the Window Toolbar. In the Annotation dialog box, enter the words Clone into pUC19 in the Annotation Text Field and click OK. Your text appears, selected, just above the center of the molecule. Drag the annotation directly under the TC(R) label (Figure 6.8). Double-click on the annotation and select Font in the Properties dialog box. Change the font to Arial 10 point bold. Click the OK buttons in both dialog boxes.
An annotation can be removed by selecting it and choosing **Edit > Delete Annotation** on the menu bar. Select **Yes** to confirm the deletion.

### 11. Experiment with Styles

Styles allow us to assign a single format to many objects in a centralized, organized fashion. They work much like styles do in word processing software.

The plasmid pBR322 has 3 CDSs; TC(R), AP(R) and ROP. On the graphical map, locate all three CDSs and their labels. In Step 5, you defined the unique style "CDS Label" that can be used to format all CDS labels. Then you modified the features and labels in the pBR322 Molecule Display window to reflect the new formatting. In Step 9, you modified the TC(R) label and now its format attributes are different from those prescribed by the style "CDS Label" style. Vector NTI permits graphics objects to be formatted differently than prescribed by their styles. This simplifies graphics map editing and allows one to change a style using the formatted object as a template.

Click on the TC(R) label and open its associated shortcut menu with a right click. The **Redefine ‘CDS Label’ Style by Example** menu item is enabled because the format attributes of the TC(R) label are different from those prescribed by its style. Click on the **Redefine…** command.

The CDS Label Style is changed and all three CDS labels automatically change format as well.

**Note:** Although the style is changed, it still has not been saved to the Display Setup Profile. See Step 12.

A simple alternative way to redefine a new formatting style for labels or features in the same display window is to use the Copy/Paste feature. Right-click on the AP(R) signal and open the Properties dialog box from the shortcut menu. Reformat the fill pattern and color on the Fill tab and click **OK**. Now select the re-formatted object and select **Copy Style** from the shortcut menu. Select the TC(R) arrow and select **Paste Style** (Figure 6.9) from a new shortcut menu. You can redefine the set style by selecting the Redefine… option on the shortcut menu.

![Figure 6.9 Using the Copy/Paste feature to redefine a new formatting style for labels](image)

You can use styles for automatic formatting of functional symbols—arrows, blocks, etc.—as well as for text. Styles for graphics objects store not only color and pattern attributes, but also geometrical characteristics such as width and arrowhead dimensions. This is a powerful feature for creating publication-quality figures.

In spite of the fact that the graphical formatting of pBR322 in the Display Window was changed, the molecule is not marked as modified (with an asterisk on the title bar). Graphical modifications of molecules are not stored in the database, but in a molecule document file. The actual
12. Save the pBR322 Molecule Display to a Molecule Document file

You can save the contents of the pBR322 Display window with its graphical formatting changes to a Molecule Document File (NOT the VNTI Database). Select File > Save As > Save as File tab.

In the dialog box, name the file `pBR322.gb` and press OK. If Vector NTI was configured for Internet publishing, a confirmation dialog box appears asking your permission to create an HTML file describing the molecule. Click Yes, and Vector NTI notifies you that an HTML file is created.

**Note:** The Molecule Document file you created is completely independent of the Vector NTI database. It contains all the molecule information and any results of molecule analyses and the graphical representation of the molecule. Think of a Molecule Document as a self-contained “snapshot” of your work with a molecule. You can exchange Molecule Documents with your colleagues, publish them on the Internet, send them through electronic mail, etc.

To close the pBR322 molecule window, select File > Close on the menu bar.

Vector NTI now detects that the style sheet associated with the pBR322 Display window has changed (remember that you redefined the “CDS Label” style) and requests confirmation to save them to the profile. Click Yes. In the Save Current Display Settings As dialog box that opens, select Blue from the Settings Name drop-down menu, confirm, and click OK. Since you already have a style sheet named Blue, choose Yes for Save Unused Styles. This overwrites the original Blue style sheet with the modified one. Sometimes this option does not appear until you try to close the molecule.

After the modified style sheet is saved, Vector NTI closes the pBR322’s Display window.

Now open the Molecule Document you created, with File > Open > Molecule File tab, and select the name of the saved document file from the file list. Your Molecule Document opens with all of your changes preserved.

**Note:** Every style required by a document is saved within the Molecule Document itself. This means that, unlike molecules opened from the VNTI Database, the pBR322 molecule document will always look the way you formatted it when you saved it to a Molecule Document File.

Note that the window title now reflects the fact that the Molecule Display window displays the molecule data from the document, NOT from the Vector NTI database. No matter how many times you modify the molecule and save it back to the Molecule Document, pBR322 in the VNTI Database remains unchanged.

**Note:** Vector NTI does not support the whole range of molecule operations on the molecules in Molecule Documents. Sometimes, Vector NTI prompts you to save the molecule into the database first and repeat the operation on the molecule saved in the database. The Demonstration version of Vector NTI has extra restrictions on the kind of operations you can perform on such “external” molecules.

13. Use the Saved Style Sheet to Format Molecule Display Windows

Now we can use the saved settings to format other molecules as well. Close the pBR322 Display window. Click the Open Document button ( ) on the main toolbar. In the Open dialog box on the Database DNAs/RNAs tab, select the M13 molecule from the molecule list and click OK. All CDS features and labels display the default settings. In the Molecule Display window, click the Display Setup button. Select the Blue style sheet from the Setup Profile list on the
drop-down menu and click **OK**. In the Molecule Display Window for the M13 molecule, note that all CDS labels now have the formatting we defined in a "CDS Label" style (Figure 6.10).

![M13 molecule diagram](image)

*Figure 6.10 Using a saved style sheet to format Molecule Display windows*

You can define and give names to several independent Display Setup settings like **Blue** and use them to reformat graphics representation of molecules in Molecule Display windows.

To switch to another style sheet, click on the **Display Setup** button in the Window toolbar, and select the name of the style sheet in the popup menu. Vector NTI reformats the active display window according to setting you choose.

14. **Close the Display Window and Exit Vector NTI**

To finish your session with Vector NTI, close the Molecule Display Window using **File > Close** on the menu bar. Exit Vector NTI using **File > Exit**.
Introduction

This tutorial introduces you to Vector NTI’s Database Explorer, a simple and flexible interface for managing and manipulating various kinds of objects in the Vector NTI database. The Explorer format makes it relatively simple to store, sort, and manage information such as molecules created in your laboratory, enzymes stored in your freezer, molecules with particular antibiotic resistance capabilities or oligonucleotides designed for PCR.

At the end of this tutorial session, you will be able to:

- Navigate the VNTI Database Explorer
- Use menu and toolbar commands
- Alternate between database subsets
- Select and edit listed subset objects
- Search the database for objects with specified features
- Create new subsets
- Add new user-fields to database objects
- Use User-Field Manager to display new user-field columns
- Use Contact Manager to manage contact information

1. Launch Vector NTI

To launch the Vector NTI Database Explorer, select Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI Explorer.

Alternatively, you can access the Database Explorer window by using the Local Database button ( ) on the Main Toolbar of the Vector NTI workspace.
Note: By default, the Vector NTI Database Explorer launches independently of the Vector NTI workspace. However, it can be configured so the Database Explorer opens automatically with the Vector NTI workspace each time a Vector NTI session is initiated. From a Molecule Display window, select Edit > Options from the menu bar. On the General tab of the Options dialog box, select Open Local Explorer At Startup.

2. Inspect the Database Explorer Window

Maximize the Vector NTI Database Explorer window. This window displays the contents of local and shared databases. The databases are organized into several tables containing subsets of the respective type. Each Database Explorer window has two panes: the Subset Pane on the left and the selected subset Database Objects Pane on the right.

Locate the database Table box in the upper left-hand corner of the screen and click on the drop-down menu arrow to view the various Database Tables. To alternate between database tables, select a table from the drop-down menu or select from the Table menu on the menu bar.

Figure 7.1 A list of protein subsets is displayed in the Subset Pane

In the Database Table drop-down menu, select the Protein Molecules table. The list of protein subsets is displayed in the Subset Pane, initially containing only one subset called Protein Molecules (MAIN) (Figure 7.1).

The subset contents are listed in the Database Objects Pane. This subset contains all of the protein molecules in the database. The contents of each of the (MAIN) default subsets are maintained by Vector NTI similarly with all of the respective database objects for that table. Scroll through the list of proteins to become familiar with the database format.

The Vector NTI Database Explorer window has a single toolbar containing the most frequently used commands. Most commands can also be initiated from the menu bar.
3. Experiment with View Options

Pull down the View menu to review the available viewing options. (Options are duplicated by toolbar buttons.) You can view your molecules as large or small icons, a compact list and a tabular form with columns showing various object properties (the “details” mode).

By default, the Database Explorer shows proteins in the details mode. Click on each of four mode buttons at the right of the toolbar to view the object list in each mode, and then switch back to the details mode.

To sort database objects in ascending or descending order, select View > Arrange Icons or, if you are in the details mode, click on the column headers. Click on the Length column header several times to sort the molecules by length. Adjust the widths of the column by click + dragging the small lines separating the column headers.

4. Experiment with Selection and Menus and Edits

The Database Explorer window combines several standard selection techniques. Select multiple objects by:

- SHIFT + CLICK to select all objects within a range
- CTRL + CLICK to select non-contiguous objects in a list
- “wire-frame” selection (press the mouse button somewhere outside item names / icons and drag the selection rectangle)

Some menus are disabled when nothing is selected. Object and subsets have shortcut menus listing various associated commands.

To rename both subsets and database objects, click on the name of the item to be renamed, wait a second, then click on the same name again. Clicking two times in rapid succession is interpreted as double-click, but waiting momentarily before the second click is interpreted as an in-place editing request. An active edit text box opens on the item name, allowing you to edit the name.

5. Learn About Subsets

In the Protein Molecules (MAIN) table, click on the Name column header (Database Objects Pane) to sort the list by name in alphabetic order. Select the 41BB_HUMAN and 41BB_MOUSE molecules and click the New Subset button ( ) on the toolbar. Vector NTI creates a new subset called Group 1. Name the new subset 41BB in the active text box and press ENTER. The new subset is selected in the left pane and its contents are listed in the right pane (Figure 7.2).

Figure 7.2 Creating a new subset

[Image: Figure 7.2 Creating a new subset]
You can easily add objects to subsets by drag-drop operations within the Database Explorer window. Other subset operations are available from the Table menu or from the subset shortcut menu.

**Note:** The difference between Vector NTI’s subsets and Vector NTI’s file system’s folders is basic. Because a subset is a set of objects grouped by a specified criteria, one object can exist in several different subsets at one time. On the other hand, one object cannot exist in several folders at the same time, but you can make a copy of an object to place it in another folder. Note that the 41BB_HUMAN and 41BB_MOUSE molecules were not moved out of the MAIN subset. You just created another group (subset) based on their common features.

### 6. Search the Database

The Database Explorer has a sophisticated searching capability enabling you to search by different criteria and logically combine search results.

Let’s find all DNA molecules containing an ampicillin resistance gene, create a new subset for them and add appropriate user fields. Open the DNA/RNA (MAIN) molecule database and press the **Search** button ( ) on the Database Explorer’s toolbar, opening the Molecule Database Search dialog box (Figure 7.3):

![Figure 7.3 Molecule Database Search dialog box](image)

Check the Features Filter option and press the **Features Setup** button. In the Search Conditions: Features dialog box, click the **Add** button, opening the Add Features dialog box (Figure 7.4) where you choose the feature for which to search. Select **CDS** in the Feature Type window. Click the **Add only features of this type...** radio button. Browse through the list in the right pane, locate the AP(R) and check its box.

![Figure 7.4 Add Features dialog box](image)
Click **OK**, closing the box and adding the AP(R) feature to the Search Conditions: Features dialog box (Figure 7.5).

![Figure 7.5 Adding a new feature to the Search Conditions: Features dialog box](image)

Click **OK** to return to the Database Search dialog box.

Make sure that all filters except the Attributes and Feature Filters are unchecked and that the “DNA/RNA (MAIN)” subset is selected in the “Look in” drop-down menu. Click the **Start Search** button. Vector NTI scans the entire DNA/RNA table, locating all molecules with an ampicillin resistance gene.

When the Insert Search Results dialog box opens to select a subset for the search results, enter a new subset name, **Ampicillin Resistance** (Figure 7.6).

![Figure 7.6 Insert Search Results dialog box with new subset name, Ampicillin Resistance](image)

Click **OK**. Vector NTI informs you of the number of items in the result set.

The subset is created (you can see it in the All Subsets Pane), but the Search dialog box is not closed. This is so that you could refine your search using the result set as an input set of a next search. Click the **Close** button to close the Search dialog box.

### 7. Examine Vector NTI's User Field Manager

Each of the columns in a subset Objects Pane is a user field displaying identifying characteristics of subset objects. A user field can contain any laboratory-specific and user-specific information necessary to provide an appropriate description of the object. For instance, you might identify all of the enzymes that are stored in your laboratory so they can easily be sorted in a
larger database. User Fields can be imported, exported, displayed, printed and managed just as the regular database fields.

Click the new Ampicillin Resistance subset to open it and view its molecules, all of which presumably contain an ampicillin resistance gene. Now let’s add the ampicillin user field to the molecules so that feature will be displayed in a user field column. That will allow you to sort molecules with that feature in the (MAIN) database.

Choose **Database > User Field Manager** from the menu bar, opening the User Fields Manager dialog box (Figure 7.7):

![User Fields Manager dialog box](image)

This dialog box allows you to define new User Fields and to perform global operations on user fields such as Edit, Rename, Delete as well as Import/Export operations. Importing and exporting allows you to exchange your list of field definitions with your colleagues.

Vector NTI maintains a local list of descriptions of “known” User Fields that are presented to the user when he or she creates new database objects. This list can be viewed and modified using the User Field Manager.

**Note:** Field definitions and field values are stored and exported with database objects. When these objects are imported, Vector NTI helps you to add the field definitions to the local User Field definitions list.

### 8. Define a New User Field

Click on the **New** button ( ) in the upper left of the User Field Manager dialog box, opening the User Field Definition dialog box.
Enter **Ampicillin Resistance** in the Name box, and using the Type drop-down menu, select **One of Predefined Strings**. In the Predefined Values text box, type **Ampicillin Resistance** and click the < Add button (Figure 7.8).

Click **OK**, closing the window. Your field description is added to the field list.

Now let's assign this value to all molecules listed in the Ampicillin Resistance subset we just created. Back in the User Field Definitions box, in the large window (Field Definition List), select the newly listed **Ampicillin Resistance**. Now in the box in the lower right corner, select [+ ] DNA/RNA Main to open the DNA/RNA subsets; select the subset **Ampicillin Resistance**. In the Global Operations area, click the **Set 'Ampicillin Resistance'** button (Figure 7.9). In the Ampicillin Resistance dialog box that opens, select the field value string **Ampicillin Resistance** and click **OK**. Click **OK** to confirm your action to insert the field value. Vector NTI performs the operation and returns you back to the User Field Manager. Press **Close**.

### 9. View the New Field in Database Explorer

Now let's customize the Database Explorer view to display the new field. Back in the DNA/RNA (MAIN) Table Explorer window, in the Database Objects Pane, right click on a column heading.
In the shortcut menu that opens, select **Columns**. The Columns dialog box allows you to control the display and order of columns in the details view of the Explorer window.

![Figure 7.10 Columns dialog box](image)

In the **Available Fields** box, select **Ampicillin Resistance** and click the **Forward** button (>>) , moving the field to the **Displayed Fields** box. Click the **Move Up** button several times to move the Ampicillin Resistance column just below the Length column (Figure 7.10). The descending order in the list is the left to right display order of the columns.

Click the **OK** button, updating the Database Explorer to display the new field as a column in the list pane. Return to the DNA/RNA subset by clicking on DNA/ RNA Molecules (MAIN) in the All Subsets Pane or by selecting **Table > DNA/RNA Molecules** from the Main menu (Figure 7.11).

![Figure 7.11 Updating the Database Explorer to display the new field as a column in the list pane](image)

Scroll through the Objects Pane, noting that Ampicillin Resistance is noted for some of the objects. Now click on the Ampicillin Resistance column heading, sorting the objects according to this feature. Click again and note the sorting changes. **While molecules in this subset can be sorted according to this new user field, as you may expect, all DNA molecules in the Ampicillin Resistance subset display “Ampicillin Resistance” in that column. You could create a column for any feature by which you wish to sort molecules using these same steps.**
Sorting by the Ampicillin Resistance column takes longer than sorting by other columns like Length. This is because the Ampicillin Resistance column is not cached in index tables and thus is slower to access. Caching requires extra memory, but it usually makes sense to turn it on for fields that are displayed as columns in the list pane.

To turn on caching of Ampicillin Resistance, open the User Field Manager dialog box again. Double-click on the Ampicillin Resistance definition to edit it, check the **Cache In Index Table** box and click **OK**. Vector NTI asks for confirmation and then rescans the database and rebuilds the database indices. Close the User Field Manager dialog box and try again to sort the DNA molecules by clicking on the Ampicillin Resistance column header. Sorting should work faster now.

### 10. Examine Vector NTI’s Contact Manager

The Contact Manager is a tool for creating, updating, deleting and organizing contact records. This list, to simplify tracking of database objects, is automatically updated when you import molecules or proteins created by your colleagues. You can enter any suitable information about yourself; this information then is attached to your molecules and is automatically added to the contact list of those colleagues who import your molecules.

Select **Database > Contact Manager**, opening the Contact Manager dialog box (Figure 7.12):

![Figure 7.12 Contact Manager dialog box](image)

This dialog box displays a list of all stored contact records. There are five operation buttons in the upper left corner of the Contact Manager allowing you to perform various operations on the contact records, including Add New, Edit, Rename and Delete.

### 11. Close the Molecule Display Windows and Exit Vector NTI

You have now taken a close look at Vector NTI’s Database Explorer capabilities. Close all Display windows and exit Vector NTI as you learned in previous tutorials.
Introduction

This chapter introduces you to tools for designing and analyzing PCR primers and procedures. At the end of this tutorial session, you will be able to:

- Design and analyze PCR primers using Vector NTI's primer design tool
- Add restriction enzymes to PCR primers
- Set parameters for optimum primer/target interactions and successful PCR
- Save primers and PCR results in the database
- Add primers to the Oligo List
- View and analyze PCR results
- Order Custom Primers from Invitrogen

Note: Press the F1 key to open Online Help for all of the dialog boxes in this tutorial.

1. Launch Vector NTI and Open a Molecule Display Window

- Launch Vector NTI using techniques described in previous tutorials (Select Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI).
- Using techniques learned in earlier tutorials, open a Molecule Display window containing ColE1.

2. Select the Region of ColE1

- Click the Graphics Pane button on the Window Toolbar to activate the Graphics Pane.
- Select Edit > Set Selection. In the Set Selection dialog box, select the 5200–6400 region of ColE1, the target fragment for PCR.
3. Prepare for PCR Analysis

To perform PCR analysis on the selected fragment, select **Analyses > Primer Design > Find PCR Primers** on the menu bar. The Find PCR Primers dialog box appears (Figure 8.1):

![Figure 8.1 Find PCR Primers dialog box](image)

With this dialog box, you can direct Vector NTI to search for convenient sense and antisense primers according to your specifications and to check primers’ uniqueness.

Click the **More>>** button on the lower right to enable a complete view of the Find Primers dialog box (Figure 8.2):

![Figure 8.2 Find Primers dialog box](image)

The **Primer** tab allows you to designate basic parameters for the PCR reaction and the primers that are generated. After clicking the **More>>** button, you can also define primers yourself and attach short sequences at the 5’ and/or 3’ ends of the product. (If you want the PCR target length to be the full extent of the selected target region, make sure that the Region of Analysis and Product Length sections of the dialog box reflect your selections.)
Note: The calculation for Tm is dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Make sure to adjust these parameters according to your reaction conditions when performing your own PCR analysis to ensure that you obtain accurate Tm values.

For this tutorial, review the options on the Primer tab, but leave all the options unchanged.

Select the Amplicon tab (Figure 8.3). The options on this tab allow you to customize parameters relating to the resulting PCR product. %GC content for the product or a portion of the product next to the primer annealing site and allowed bases adjacent to the primer annealing site can be specified.

Review the options, but leave all the options unchanged.

Select the Structure tab (Figure 8.4). The options on this tab set acceptable limits for nucleotide repeats, palindromes and hairpin loops for the primers. You can also check your primers/product for a selected group of restriction sites from this tab.

Inspect all options, leaving them unchanged.
Select the **Pairs** tab (Figure 8.5). Options on this tab specify how closely parameters such as Tm and %GC, etc. must match between two primers in a generated primer set.

Inspect all options, leaving them unchanged.

Select the **Similarity** tab (Figure 8.6). The options on this tab determine the similarity relationship between the primers and the target sequence.

Inspect all options, leaving them unchanged.
Select the **3' end** tab (Figure 8.7). The options on this tab allow you to set specifications for the 3’ end of the primers generated by Vector NTI. Parameters such as dG and specific nucleotide content for the 3’ end of both sense and antisense primers can be set here.

![Figure 8.7 The 3’ end tab allows you to set specifications for the 3’ end of the primers generated by VNTI](image)

Inspect all options, leaving them unchanged.

Select the **Uniqueness** tab (Figure 8.8). The options on this tab determine the uniqueness of the primers generated with respect to the PCR product. These parameters can be used to help ensure that generated primers bind to the desired template area with greater specificity than to the rest of the PCR product.

![Figure 8.8 The Uniqueness tab determines the uniqueness of the primers generated with respect to the PCR product](image)

Inspect all options, leaving them unchanged.

Select the **Qualities** tab (Figure 8.9). These parameters govern primer quality by determining how much weight should be assigned to parameters specified on the other tabs of the Find
Primers dialog box. These values affect scoring functions that evaluate the quality rating of the primer sets generated.

Inspect all options, leaving them unchanged.

Select the Filters tab (Figure 8.10). On this tab, you can select features that should either be excluded or included as regions to be considered for primer design. For example, if your sequence contains Repeat features, you may want to exclude them to obtain primers with greater specificity. Features can be added or removed by pressing the + and – buttons.

Leave all the options unchanged as you inspect them.

Notice the Load and Save buttons on the lower left of each tab in the Find Primers dialog box. These buttons allow you to save your PCR settings to a file and load the settings file in for subsequent analyses. This precludes having to reset the various parameters manually for frequently used PCR condition settings. You need only save/load the settings in one of the tabs and the parameters for all the tabs in the Find Primers dialog box will be saved/loaded.
4. Perform PCR Analysis

In the Find Primers dialog box, press the **OK** button in the lower center, initiating the PCR analysis. When it is finished, a new folder is created, added and opened in the Text Pane of the ColE1 Display window.

5. Inspect the PCR Analysis Folder

Move the split bars as necessary to optimize viewing in the Text Pane. In the open PCR Analysis folder, inspect the subfolders.

- Each subfolder describes one possible set of primers for amplifying a fragment of the length within the maximum and minimum lengths specified in the Find Primers dialog box.
- The subfolders are listed in order of decreasing primer set quality rating which is calculated based on the importance factors assigned in the **Qualities** tab in the Find Primers dialog box. (Generally, maximum rating = 171.)
- Each subfolder contains useful information for the PCR reaction such as product melting temperature, optimum annealing temperature, sense and antisense primers with their melting temperatures, free energies, etc.

The sense and antisense primers are indicated with the symbol "[ ]" followed by their nucleotide sequences. (In Vector NTI, this same symbol is used to mark oligonucleotides created by many of the program’s functions such as PCR analysis, molecule design, and hybridization. You can perform oligo analysis on any nucleotide sequence marked by this symbol in the Text Pane of a Molecule Display window.)

6. Perform Oligo Analysis on a PCR Sense Primer

To perform oligo analysis on the sense primer given in the first subfolder of the PCR Analysis folder, scroll through the Text Pane and verify that the folder labeled “**#1: Product of length 901**” is open. (Open it by double-clicking it, if necessary.)

Inside this subfolder, position the cursor on top of the sense primer, represented by the line **ATTACTCTTTTGAATGGTACTCCTG**. Open the corresponding shortcut menu by right clicking on the primer sequence. Choose the **Analyze** command, opening the Oligo Analysis dialog box (Figure 8.11):

![Figure 8.11 Oligo Analysis dialog box](image)
In this dialog box you can set numerous parameters for oligonucleotide analysis. Leave the default parameters unchanged and press the **Analyze** button.

**Note:** The calculation for Tm is dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Make sure to adjust these parameters according to your reaction conditions when performing your own oligo analysis to ensure that you obtain accurate Tm values.

The analysis results appear in the dialog box. Note the biochemical and structural parameters in the Results column. Press the **Dimers & Hairpin Loops** button to inspect the dimers and hairpin loops that may form if this primer is used.

Press the F1 key to view Vector NTI’s on-line help information for this dialog box. Press the **Close** button.

### 7. Order Custom Primers

In Vector NTI Advance 10, you can order custom primers that you have generated and analyzed in the software. Custom primers are synthetic oligonucleotides made with your specified sequence for use in a variety of applications from PCR and sequencing to probes for gene detection. Custom modifications, such as standard deoxynucleotides, modified bases, 5’ modified nucleotides (including fluorescent dyes, enzyme conjugates), and S-oligos for antisense studies can be designated. Specifications also include scale of synthesis, level of purity and 5’ or 3’ modifications.

To initiate the ordering process, inside the PCR Analysis folder, scroll through the Text Pane and verify that the folder labeled “**#1: Product of length 901**” inside the subfolder is selected by positioning the cursor on top of the sense primer, represented by the line ATTACTTTTGAATGGTACTCCTG. Open the corresponding shortcut menu by right clicking on the primer sequence. Choose the **Add to Oligo List** command, opening the New Oligo dialog box, with the **General** tab activated. Enter **Sense primer - ColE1** in the name box and click **OK**. Vector NTI verifies the oligo has been added.

Open the Oligo List dialog box by selecting **List > Oligo List**. Select the oligo named **Sense primer - ColE1** and click the **Order** button.

![Figure 8.12 Oligo List dialog box](image)

The Invitrogen webpage opens automatically, and your primer sequence is entered in the custom primer order form. If you in reality wanted to order the primers, at this point you would proceed to checkout. For this tutorial, however, simply inspect the Invitrogen Custom Order webpage, then exit and return to the Vector NTI workspace.
8. Save a PCR Primer to the Database

Now let's save the same primer to the database for future use. Right-click on the ATTACTCTTTTGAATGGTACTCCTG line in the Text Pane, to open the corresponding shortcut menu and select Save To Database. This opens the New Oligo dialog box, with the General tab activated (Figure 8.13).

![New Oligo dialog box with the General tab activated](image.png)

Enter **Sense primer - ColE1** in the name box and click on the Oligo tab. The oligo sequence and description appear automatically in their respective text boxes.

Click on the Keywords tab. Click in the New Keyword entry box and enter **PCR-ColE1**. Press the < Add button to add your new keyword to the keyword list. Then select the existing SENSE_PRIMER keyword and add it to the keywords list, too.

Press the OK button. The primer is saved to the database under the name **Sense primer - ColE1**.

9. Add Restriction Enzyme Sites to the PCR Primers

You will load the sense primer you just saved to the database and then attach sequences to the 5' and 3' ends of the product to allow convenient cloning. Vector NTI will then find an optimum antisense primer.

Activate the Graphics Pane again and make sure the 5200–6400 region is still selected. (Check the status bar.) Open the Find Primers dialog box again by selecting Analyses > Primer Design > Find PCR Primers on the menu bar.

Locate the User-Defined Primers box in the Find Primers dialog box by pressing the More>> button on the Primer tab. (Any previously selected restriction sites will appear in the User-defined Primers Sense and Antisense text boxes by default. To continue with this tutorial, select each and press the DELETE button so each text box is blank.) Press the Browse button ( ) to the right of the Sense Primer box. In the dialog box that appears, showing oligonucleotides saved in the database, select the “Sense primer - ColE1” oligo you just saved and click OK. The “Sense primer - ColE1” oligo is loaded into the Sense Primer text box (Figure 8.14). Note that you are not specifying the anti-sense primer, but leaving that up to Vector NTI.
To attach a *BamH*I cloning site to the 5’ end of your sense primer, locate the Attach to 5’ Terminus of Sense Primer box and press its **Browse** button. In the Choose Database Enzyme dialog box that opens, select the MAIN Enzymes subset. Scroll down to find and select *BamH*I. Click **OK**. The *BamH*I recognition site is entered into the Attach to 5’ Terminus of Sense Primer box.

To attach a *HindIII* cloning site to the 5’ end of the antisense primer, locate the Attach to 5’ Terminus of Antisense Primer box and press its **Browse** button. In the dialog box, select the MAIN Enzyme subset and *HindIII* and click **OK**. The *HindIII* recognition site is entered into the Attach to 5’ Terminus of Antisense Primer box.

Figure 8.14 The “Sense primer - ColE1” oligo is loaded into the Sense Primer text box; restriction enzymes are designated for the sense primer terminus.

Locate and check the Check Primers for Restriction Sites From check box on the **Structure** tab, directing Vector NTI to search for certain restriction sites in the PCR product-plus-primer fragment you are creating. The restriction sites searched for are listed in the Restriction Endonuclease subset shown in the Check Primers for Restriction Sites From drop-down menu. Choose the **Palindromes/Non-Ambiguous** subset from the list (Figure 8.15).

Figure 8.15 Searching for certain restriction sites in the PCR product-plus-primer fragment on the Structure tab.
Now press **OK** at the lower center of the Find Primers dialog box. Press **OK** when Vector NTI shows the warning dialog box. Click **OK** to override the previous PCR analysis results.

A new PCR analysis is performed using the additional information you have entered.

**10. Inspect the New PCR Analysis Results**

Inspect the subfolders of the new PCR Analysis results in the Text Pane.

Note that this time, all of the new options use the same sense primer, “Sense primer - ColE1”, and that the REN sequences have been attached at the 5’ ends of both primers. (They are separated from the primers by a space.) Only the antisense primers vary because they were not defined in the last dialog box.

The RENs appearing beneath each primer indicate those from the selected Palindromes/Non-Ambiguous subset whose recognition sequences are found within the generated fragment. The number of times that a REN recognition sequence is found within the PCR product is indicated. If no number is given after a REN’s name, then that REN is found only in the primer or the attached sequence and not within the product.

**11. Save the PCR Product to the Database and Open a Display Window**

Place the mouse cursor on the folder labeled “#1: Product of length 919,” and open the shortcut menu with a right click. Choose the **Save to Database and Create Window** command. In the dialog box that opens, in the Name field, enter **PCR Product 1**, and click **OK**. Choose the Main subset and click **OK**. The molecule is saved to the database, and a new Molecule Display Window is created containing the molecule.

**12. Arrange the View of the Product**

Arrange the new Molecule Display window to show your PCR product conveniently. Activate the Text Pane, and click the **Link Panes** button ( ) on the Window Toolbar. Most of the symbols in the Graphics Pane disappear because most of the folders in the Text Pane are closed.

Place the cursor on the Feature map folder and open the shortcut menu. Choose the **Expand Branch** button ( ), opening the Feature map folder and all its subfolders. Now the signals are displayed in the Graphics Pane.

Locate the Restriction Map folder and open it with a double-click. Open the BamHI and HindIII folders by double-clicking them. These sites now appear in the Graphics Pane.

Activate the Graphics Pane and press the **Standard Arrangement** button. Arrange the graphics to fit the window so that the restriction sites and primers can be seen. (Use the **SHIFT + Zoom In** ( ) and **Zoom Out** ( ) buttons.) Use the **CTRL + Zoom** buttons to change only the horizon-
tal dimensions of the graphical map. (You can enhance your picture further using picture editing techniques described in Chapter 6.)

![Figure 8.16 Standard Arrangement of restriction site and functional signal labels](image)

13. Close Display Windows and Exit Vector NTI

This concludes the PCR analysis tutorial. Close all Molecule Display windows using the techniques described in previous tutorials: File > Close, then File > Exit.
Introduction

This chapter introduces you to creation of a new DNA molecule by Construction. In Vector NTI terminology, Construction means creating a DNA molecule from fragments that are completely defined and made compatible by the user.

At the end of this tutorial session, you will be able to:

- Use Fragment Wizard to define the needed fragments for the Goal Molecule Definition List
- Name and describe the new molecule in the Construct Molecule dialog box
- Verify that the fragments are correctly defined
- Direct Vector NTI to construct your new molecule
- Modify the fragment termini biochemically, when necessary, to allow completion of the Construction process

DNA molecules can be composed of fragments of existing DNA molecules, linkers, adapters, dummy fragments and so forth. Most of the fragments you will use to create new molecules are fragments of existing molecules, and most of the work involved in creating a new molecule is in defining them. Fragments like linkers and adapters are much easier to describe.

The Fragment Wizard guides you through the process of describing a new molecule fragment. In addition to using the Fragment Wizard, you can define fragments using the Fragment Editors available from the Construct/Design Molecule dialog box. While they can describe fragments of all types, Fragment Editors are most convenient for defining linkers, adapters, and so forth.

Molecules that you or Vector NTI create from fragments of existing molecules are called constructed molecules. Molecules that you import or describe by hand are called basic molecules because they enter the database as complete units rather than being built in Vector NTI from fragments.

For more information regarding other methods of adding new molecules to Vector NTI's database, such as design, importing molecules, assembling a molecule from a sequence text file, refer to Chapters 10 and 11.
1. Launch Vector NTI

Launch Vector NTI using techniques you learned in previous tutorials.

2. Create Display Windows for pBR322 and pUC19

Using the techniques you learned in previous tutorials, click the Open button and open pBR322. Then click the Open button again to open pUC19. Two display windows are now open, one for each plasmid. Both molecules, pUC19 and pBR322, can be viewed by selecting Window > Cascade from the Main menu (Figure 9.1), or you can move from one molecule to the other by selecting the molecule in the drop-down Window menu.

![Figure 9.1 Opening and Viewing two Display Windows using the Cascade option in the Window menu.](image)

3. Arrange the Display Windows

Expand the working space for the Graphics and Sequence Panes by reducing the Text Pane in each Display window. Adjust the graphical map for each plasmid so you can easily read the features and labels.

4. Define the First Fragment (from pUC19)

The first fragment of the goal molecule includes most of pUC19, with the 5’ end of the fragment being the Smal restriction site and the 3’ end being the EcoRI restriction site. The second fragment is from pBR322, the 5’ end being the EcoRI site and the 3’ terminus being the AvaI restriction site.

Activate the Graphics Pane of the pUC 19 Display window. Click the Add Fragment to Goal List button ( ) to open the Fragment Wizard (Figure 9.2).
This dialog box guides you through the process of fragment selection, with several sequential screens describing the corresponding actions you should take. If you make a mistake you can return to the previous step using the Back button.

- **First screen** (Figure 9.2): Select the Construction Fragment option. Leave the Insert Inverted option unchecked, and click the Next button to proceed.

- **Second screen** (Figure 9.3): Drag the dialog box out of the way (left-click with the cursor on the blue title bar) so you can see the Smal and EcoRI restriction sites on the molecule graphical map. To determine the 5’ terminus of the new fragment, click on the Smal restriction site label in the Graphics Pane. The Set to a Restriction Site option is auto-
matically checked and the name and position of the restriction site is displayed in the Fragment Wizard. Click the Next button in the dialog box to proceed.

![Figure 9.4 The third screen of the Fragment Wizard determines the 3’ terminus](image)

- **Third screen** (Figure 9.4): To define the 3’ terminus of a new fragment, hold the SHIFT key and click on the **EcoRI** restriction site label in the Graphics Pane. The **Set to a Restriction Site** option is automatically checked and the name and position of the 3’ end are displayed.

  **Note:** If you do not hold down the SHIFT key when selecting the 3’ end, Fragment Wizard reverts to choosing the 5’ site.

  Both labels in the Graphics Pane are now highlighted. The selected fragment is indicated by the wire-frame in the Graphics Pane. Click the Finish button in this screen to complete the definition of the fragment.

![Figure 9.5 Adding the first fragment to the Molecule Goal list](image)

Check the description of the fragment in the New Fragment message box: 5’ **SmaI**, and 3’ **EcoRI**. If there is an error, press the Cancel button and return back to the Fragment Wizard. Otherwise, press the Add to List button, adding the first fragment to the Molecule Goal list (Figure 9.5).

**5. Define the Second Fragment (from pBR322)**

On the toolbar, select **Window > pBR322**, which opens the corresponding Molecule Display window. Activate the Graphics Pane. Click the **Add Fragment to Goal List** button ( ) to display the Fragment Wizard dialog box again. Move the dialog box so that you can see the **EcoRI** and the **AvaI** restriction sites on the graphical map.
- **First screen:** Select the **Construction Fragment** option. Leave the Insert Inverted option unchecked, and click the **Next** button to proceed.

- **Second screen:** To determine the 5' end of the new fragment, click on the *Eco*RI restriction site label in the Graphics Pane. The **Set to a Restriction Site** option is automatically checked, and the name and position of the restriction site are displayed. Click the **Next** button to proceed.

- **Third screen:** To define the 3' terminus of the new fragment, hold down the **SHIFT** key and click on the *Ava*I restriction site label in the Graphics Pane. The **Set to a Restriction Site** option is automatically checked and the name and position of the restriction site are displayed.

Click the **Finish** button to complete the definition of the fragment.

In the New Fragment message box, verify the fragment definition: 5' *Eco*RI and 3'*Ava*I. If the description of the fragment is correct, press the **Add to List** button. The second fragment is added to the Molecule Goal list.

6. **Inspect the Goal List**

To verify that the fragments in the Goal Molecule Definition List are correct, click the **Open Goal List** button ( ) on the Main Toolbar. This opens the Lists dialog box (Figure 9.6), where the two fragments are listed in the Fragments section.

![Figure 9.6 Lists dialog box](image)

7. **Enter General Information About the New Molecule**

Click the **Run** button. Click in the Name field text box in the Construct Molecule dialog box (Figure 9.7) and enter the name of the new molecule, **TUTORIAL1**.

![Figure 9.7 Construct Molecule dialog box](image)

The radio buttons on the bottom describe the choices for the new molecule start. The first fragment in the list of component fragments is always considered the "recipient" molecule. Click the **Recipient's Start** button to position the start of the new molecule at the same place where *pUC19* starts. (If you prefer, you can choose any nucleotide of any component fragment as the
starting coordinate of the new molecule by checking the **Position ... of Fragment # ...** button. By default, it’s the first nucleotide of the first fragment.

Click the **General Info** button, opening the General Data dialog box where you can specify general information about the new molecule (Figure 9.8).

![General Data dialog box](image)

*Figure 9.8 General Data dialog box, where you can specify general information about the new molecule*

- Click in the Description field and type **Tutorial molecule #1**.
- In the Extra-Chromosome Replication field, check **Bacteria**.
- In the Replicon Type field, select **Plasmid**.
- In the text entry box of the Key Word area, enter your last name. Press the **<< Add** button to add your last name to the list of keywords for this molecule. **Keywords are a convenient way to find molecules you created—just search Vector NTI’s database for all molecules with your last name as a keyword.**
- Click **OK** to return to the Construct Molecule dialog box.

### 8. Attempt to Construct the New Molecule

Click the **Construct** button in the upper right corner of the Construct Molecule dialog box. In the Insert Molecule into Main subset dialog box, name your subset **Tutorial** and confirm the creation of the subset. Click **OK**. Vector NTI attempts to construct your new molecule and save it into the database. Vector NTI soon informs you that the left end of fragment #1 (the *Sma*I site) is incompatible with the right end of fragment #2 (the *Ava*I site).

Vector NTI cannot process your molecule. You must figure out why your fragments are incompatible and modify them appropriately so they are compatible for ligation. If you check the fragment termini, you can see the problem. The blunt 5’ end of the pUC19 fragment cannot link with the cohesive 3’ end of the pBR322 fragment. You need to modify these termini to make them compatible.

Click **OK** in message dialog boxes to return to the Construct Molecule dialog box.
9. Fill in the \textit{Ava} Site of the pBR322 Fragment to Make it Blunt

Close the Construct Molecule dialog box to return to the Lists dialog box. In the Fragments section, click on the second fragment in the list ("Molfragment of pBR322"). As you make the selection, the \textbf{Edit} button becomes enabled (Figure 9.9).

![Figure 9.9 Enabling the Edit button in the Lists dialog box](image)

Click the \textbf{Edit} button, opening the Fragment Editor dialog box (Figure 9.10).

![Figure 9.10 Fragment Editor dialog box](image)

Click on the \textbf{Right Terminus} button, because you need to modify the \textit{Ava} site that makes up this fragment’s right terminus. The Terminus Editor dialog box opens (Figure 9.11).

![Figure 9.11 Terminus Editor dialog box](image)
Because you need to perform a biochemical “operation” on the ends of the fragment, in the Terminus Editor, look in the Biochemical Operations section in the lower left corner. Although Vector NTI can handle up to three sequential biochemical operations per terminus, you need to perform only one, filling in the cohesive Aval site. Click in the first biochemical operations box. From the list of allowable options, choose **Completely Filled In**. Then click **OK** in the Terminus Editor and **OK** in the Fragment Editor, returning you to the List dialog box.

Click the **Run** button on the Lists dialog box to launch the Construct Molecule dialog box. Redefine the molecule’s name, Tutorial 1, if necessary. Press the **Construct** button. Select the “Tutorial” subset in the Insert DNA/RNA Molecule(s) into Subset dialog box and click **OK**. In the new dialog box, click the **Overwrite** button to overwrite your previously saved TUTORIAL1. Vector NTI analyzes your molecule and this time enters it into the database. Close the Lists dialog box.

### 10. Inspect Your New Molecule

After creating the new molecule and saving it to the database, a new Molecule Display window opens with your new molecule. Find TC(R) on TUTORIAL1’s graphical map. The molecule’s functional and restriction maps are created automatically. **Vector NTI is able to create the maps and nucleotide sequences for new molecules with complete accuracy, even in the most complicated cloning situations.**

![Figure 9.12 Inspecting the two fragments used to create your new molecule](image)

In the Text Pane, open the folder called “Component Fragments”. The two fragments that you used to make this molecule are listed (Figure 9.12). Their subfolders describe the left and right termini of each fragment.

**Congratulations, you have created your first new molecule with Vector NTI!** You defined all of the necessary restriction sites and biochemical operations to clone a fragment from pBR322 into pUC19 in the user-defined process called **Construction**. In Vector NTI’s **Design** mode, described in the next chapter, the program itself determines the recombinant strategy.

### 11. Close the Display Windows and Exit Vector NTI

To finish your session with Vector NTI, close the Molecule Display windows and exit the program using the methods you learned in previous tutorials.
Introduction

This chapter introduces you to the creation of a new molecule by Design, Vector NTI’s term for using its built-in biological knowledge for new molecule creation. In Design, the user defines the molecule fragments, but the restriction sites, methods of terminus modification, etc., are left up to Vector NTI. The Design process results in a construction plan that takes advantage of the best possible restriction sites and recombinant strategy.

At the end of this tutorial session, you will be able to:

- Define the recipient and donor fragments for Design, using the Fragment Wizard
- Confirm the correct order of recipient and donor fragments in the Design Molecule dialog box
- Name and describe the new molecule
- Specify Design parameters and preferences
- Direct Vector NTI to design the new molecule

Fragments are described as they were for molecule construction. Fragments for Design, however, must consist of exactly one recipient fragment from the parent molecule, which must be listed first in the Goal Molecule Definition List, and one or more donor fragments. Every nucleotide in the designated recipient fragment is included in the final molecule. Vector NTI searches the parent molecule nucleotides outside of the recipient fragment to try to find convenient restriction sites for inserting donor fragment(s).

You may want to indicate a specific restriction site at one or both ends of the recipient fragment. If a terminus of the recipient fragment is a restriction site, then Vector NTI automatically uses that restriction site in building the new molecule. You may request to save or lose such a specific restriction site. If it is not specified that the site should be retained, then that site may disappear from the final created molecule even if you did not specify that the site must be lost.

Fragments other than that listed first on the list are called donor fragments. Because the purpose of donor fragments is to carry functional signals into the recipient, you should describe a donor fragment, not by its termini, but by the functional signals it contains. Thus when you define
a donor fragment in a Molecule Display window, you should click on specific functional signals. Every nucleotide of these specific functional signals is included in the goal molecule, but nucleotides outside these functional signals may or may not be included in the goal molecule. Without at least one functional signal, a fragment cannot be treated as a donor fragment.

A donor fragment can contain flanking regions outside the functional signals. In describing a donor fragment, if you drag the edges of the selection zone beyond the functional signals you have targeted, the fragment will be defined as the functional signal(s) plus flanking nucleotides. You have defined the maximum allowable flanking regions. Vector NTI tries to reduce the flanking regions of donor fragments to be as small as possible, but does not go outside of these flanking regions when looking for restriction sites.

In this exercise, you will perform essentially the same cloning procedure as you did in the Construction tutorial. This time, however, you will simply define the fragments and allow Vector NTI to design the rest.

Follow the steps of the tutorial in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch Vector NTI and Create Display Windows for pBR322 and pUC19

Launch Vector NTI using techniques you learned in previous tutorials. Open Molecule Display windows containing the molecules pBR322 and pUC19.

2. Arrange Display Windows

In order to view both windows, select Window > Tile Vertical. The display screen for both molecules appears at the same time, making it easy to work with them side by side (Figure 10.1). Activate either window (indicated by a dark blue title bar) with a left click. Using the split bars, reduce the Text and Sequence Panes for each molecule to optimize viewing the molecule maps. Use the SHIFT + Zoom buttons to arrange the panes so you can easily read the features and labels.

3. Define the Recipient Fragment (from pUC19)

For the goal molecule, the recipient fragment is most of the pUC19 molecule. (Molecule pBR322 provides the donor fragment.) In the pUC19 Display window, activate the Graphics Pane. Click
the Add Fragment to Goal List button ( ), opening the Fragment Wizard. You learned in the previous tutorial that Fragment Wizard guides you through the process of fragment selection.

- **First screen:** Select the Design Recipient Fragment option and click Next.
- **Second screen:** For the 5' terminus of the new fragment, in the Set To a Position option, enter 500 in the text box. Click Next.
- **Third screen:** Enter 250 in the Set to a Position text box to define the 3' terminus. Click the Finish button to complete the definition of the recipient fragment.

![Figure 10.2 New Fragment message box](image)

In the New Fragment message box, check the description of the fragment (Figure 10.2). If correct, click the Add to List button, adding the recipient fragment to the Molecule Goal list.

In the Graphics Pane, the recipient fragment is marked with a wire-frame. Note that the polylinker near the 2 o'clock position is outside the selection. If you included the polylinker in the selection, Vector NTI would not be able to use those restriction sites because every nucleotide included within the selection zone in the recipient fragment will be included in the final molecule.

4. Define the Donor Fragment (from pBR322)

In the pBR322 Display window, activate the Graphics Pane, then click the Add Fragment to Goal List button ( ) opening Fragment Wizard again.

- **First screen:** Select the Design Donor Fragment radio button. Click Next.
- **Second screen:** Move the cursor until it is over the TC(R) arrow or label in the Graphics Pane. Drag the Fragment Wizard out of the way, if necessary. The cursor changes to a hand ( ), indicating that you are pointing directly at a functional signal. Click once to select the entire TC(R) signal. The selection wire-frame appears, aligned directly with the start and end of TC(R). The name of the signal is displayed in the Fragment Wizard. Click the Finish button. The default values are used to complete the definition of the fragment. Continuing with the Fragment Wizard allows specifications for more complicated design, to be covered in the next chapter.

In the New Fragment message box, check the donor fragment description. Note that the 3' and 5' termini of the donor fragment are described (correctly) as “Undefined”. Click the Add to List button, adding the donor fragment to the Molecule Goal list.

5. Inspect the Goal Molecule Definition List

Click the Open Goal List button ( ) in the main toolbar, opening the Lists dialog box. Notice that the Design button is selected, indicating that Vector NTI already has recognized that your fragments are for Design and not Construction.

In the Goal Molecule Definition List appearing in the Fragments section of the dialog box, inspect the two fragments you defined. For the recipient fragment (pUC 19, listed first), the termini are defined by nucleotide positions. For the donor fragment (pBR322, listed second), the termini are not defined (NODEF), but the fragment must contain the TC(R) descriptor.
Note: For Design to proceed correctly, recipient and donor fragments must be listed correctly. If one is listed incorrectly, select it and use the Up and Down buttons to reposition it.

6. Enter General Information for Your New Molecule

Click the Run button. Although this box is almost identical to the Construct Molecule dialog box, in the previous chapter, its name and operational mode have changed to Design Molecule. This is a result of the molecule building process you designated in the Fragment Wizard.

In the Name text box, enter the name of the molecule, TUTORIAL2. Click the General Info button. In the General Data dialog box, in the Description field, enter Tutorial molecule #2.

Set the Extra-Chromosome Replication type to Bacteria and the Replicon Type to Plasmid. To enter your last name as a keyword, select it from the existing keyword list in the drop-down menu and click the << Add button.

Click OK to return to the Design Molecule dialog box (Figure 10.3).

Check the Recipient’s Start button in the radio button group on the bottom of the Design Molecule dialog box to position the start of the new molecule at the same place where the recipient molecule (pUC19) starts.

7. Prepare to Design the New Molecule

Click the Design button in the upper right corner of the Design Molecule dialog box. When asked, select the “Tutorial” subset you created in the last tutorial, and click OK to continue. The Design Parameters dialog box opens (Figure 10.4).

Here you can select the subsets of restriction endonucleases (REns) that Vector NTI considers when designing the new molecule. For example, you could create a subset listing the REns you
89

have on hand in the laboratory and tell Vector NTI to design the new molecule using only those enzymes.

You can also indicate the transformation systems to be used in your experiments, as well as the presence or lack of extra-chromosome replication capability of your molecule in those transformation systems. You can permit or forbid the use of dephosphorylation as a pre-selection method, etc.

Make sure that the Palindromes/Non-Ambiguous REN subset is selected. For this tutorial, leave all other parameters at their default values.

8. Set Your Preferences for Molecule Design

Now click on the Preferences button, opening the Design Preferences dialog box (Figure 10.5). In this box, you can choose the parameters you prefer to create new molecules. You can designate which genetic engineering techniques are permissible and assign priorities for use of the permissible techniques. You can set preferences in the categories of fragment isolation, fragment ligation, and terminus modification.

You may find it useful for Vector NTI to design the same goal molecule several times, changing the design preferences each time. This gives you several alternatives for constructing your molecule.

Let's change the preferences for fragment ligation. Turn off the check box for blunt-blunt ligation. In a blunt-blunt ligation, both the donor and the recipient have only blunt termini. Since we are deactivating this option, Vector NTI will make sure that all fragments have at least one cohesive end.

Below the check boxes are priority lists specifying which of the techniques are preferable. Leave these at their default values.

With the Advanced Preferences, accessible through the Design Preferences dialog box, you could alter the way Vector NTI's evaluates possible design paths. For this tutorial, ignore the Advanced Design Settings.

Click the OK button, accepting the Design Preferences and returning to the Design Parameters dialog box.
9. Design the New Molecule

Click the Start Design button. Vector NTI generates many possible ways of cloning the donor into the recipient and seeks the best solution using its biological knowledge and the preferences you indicated. An optimum cloning method is quickly found and the goal molecule is created based on that best option.

10. Inspect the New Molecule

When the design is complete, Vector NTI opens a new Molecule Display window containing the molecule you have created (Figure 10.6).

![Molecule Display](image)

*Figure 10.6 Inspecting the new molecule you have created*

Maximize and inspect the graphical map and text description of your new molecule. In the Text Pane, the new Design Description folder contains instructions for creating TUTORIAL 2 at the laboratory bench.

11. Inspect the Design Plan

To verify the restriction enzymes used in the design process, they must be specified in the restriction enzyme display list. This is accomplished by clicking the Display Setup ( ) button on the Pane Selection toolbar.

In the Molecule Display Setup dialog box, check the Restriction Map checkbox and press the RMap Setup button to specify the restriction sites to be displayed. The Restriction Map Setup dialog box displays in the Use Enzymes column the restriction enzymes, whose sites are shown by default. To include the enzymes from the design process in the display list, click the < Add button.

In the Choose Database Enzymes dialog box, with the Palindromes/Non-ambiguous subset selected in the Look in: pull-down menu at the top of the dialog box, select CfrJ4I and Ball from the list of available restriction enzymes. Click OK. For more information, See Restriction Map Setup on page 219.

In the Text Pane, open the Design Description folder and its subfolder labeled Step #1.

- Vector NTI uses the CfrJ4I and EcoRI sites of the recipient, and the EcoRI and Ball sites of the donor. The donor fragment has short flanking regions on either side of the func-
tional signal TC(R). The sites chosen are compatible, so no biochemical operations are necessary to modify the termini.

- With the selected restriction sites, each fragment has one cohesive end and one blunt end, so blunt-blunt ligation is avoided as requested. The selected cloning option gives the required orientation of the cloned fragment in the recipient.

- One of the recipient’s restriction sites—CfrJ4I—is lost after ligation. This allows pre-selection of ligated molecules using CfrJ4I before transformation. Because the recipient’s CfrJ4I site is lost in the successful cloning process, this prescreening eliminates improperly ligated fragments. As only one transformation system (bacteria) was designated, Vector NTI suggests no alternatives. If you had permitted several transformation systems, Vector NTI would have chosen the system that involves the least effort.

- For clone analysis after transformation, Vector NTI suggests using the restriction enzyme Aaai. Its recognition site is new in the recombinant molecule and does not exist in the recipient.

- For alternative methods of clone analysis, the system recommends an oligonucleotide for colony hybridization and PCR primers to use in that technique.

- Vector NTI lists restriction sites close to the cloned fragment that can be used to isolate the cloned fragment from the recombinant molecule: EcoRI and BspAAI. Vector NTI has only considered the restriction endonucleases in the REN subset selected in the Design Parameters dialog box.

- The Design Description folder shows what sites the recombinant lacks, and lists all unique sites on the recombinant, divided into two groups, those inside the cloned fragment and those outside.

If the design were to consist of more than one cloning step, similar information would be given in a separate folder for each stage of the process.

12. Print Out the Design Plan

To print the design plan generated for your new molecule, activate the Text Pane of the TUTORIAL2 Display window. Close all of the folders, then open the Design Description folder and the Step #1 subfolder. Click the Print Active Pane button ( ) on the Window Toolbar, and the open folders in the Text Pane are printed to your printer. You now have a hard copy of the Design Plan for TUTORIAL 2.
INTRODUCTION

In the previous two tutorials, you were introduced to simple molecule creation using Vector NTI’s Construction and Design tools. This chapter introduces you to more complex Design procedures. In the first scenario, you will assign some complex conditions to the recipient molecule, leaving the donor fragment relatively simple. In the second, you will use a relatively simple recipient but make the donor fragment more complex. Since you are probably getting tired of pBR322 and pUC19, let’s change molecules. You will now use BPV1 and SV40.

At the end of this tutorial session, you will be able to:

- Define the recipient and donor fragments for Design using Fragment Wizard, placing complex requirements upon the recipient.
- Define the recipient and donor fragments for Design using Fragment Wizard, placing complex requirements upon the donor.
- Describe the new molecule.
- Design the new molecule.

Follow the steps in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch Vector NTI, Open and Arrange Display Windows

Launch Vector NTI using techniques you learned in previous tutorials.

Using techniques you learned in previous tutorials, open Molecule Display windows for molecules BPV1 and SV40. Select Window > Tile Vertical to work with the molecules side by side. Since you will be working with the molecules’ graphical maps, arrange the display windows conveniently.

First Design: Complicated Recipient

In the first complex design, you will insert SV40’s LARGE_T gene into the second ApaLI site of BPV1. You will direct Vector NTI to save the 5’ ApaLI site and prohibit blunt-blunt fragments. If
the donor has ApaLI sites appropriate for cutting the cloned fragment, then the problem will be simple. If not, the system must take a more complicated approach to perform the insertion.

2. Define the Recipient Fragment

In the BPV1 Display window, activate the Graphics Pane. Open the Fragment Wizard using the Add Fragment to Goal List button ( ) as you did in the previous tutorials, and follow these steps:

- **First screen:** Select the Design Recipient Fragment option. Click the Next button.
- **Second screen:** To define 5’ terminus, click on the label of the ApaLI restriction site #2 in the Graphics Pane (nucleotide 7631, or about the 11 o’clock position). Click the Next button to continue.
- **Third screen:** Select Save Site to save the ApaLI site on 5’ end of the recipient fragment. Click the Next button.
- **Fourth screen:** To define the 3’ terminus, press SHIFT + CLICK on the same ApaLI site. The name and the position of the site are now displayed in the Fragment Wizard. Click the Finish button to complete the definition of the recipient fragment.

![Figure 11.1 New Fragment message box, with 3’ and 5’ termini of the recipient fragment set on the same ApaLI site](image)

In the New Fragment message box, inspect the recipient fragment. Note that the 3’ and 5’ termini of the recipient fragment are set on the same ApaLI site (Figure 11.1). Click the Add to List button, adding the recipient fragment to the Molecule Goal list.

3. Define the Donor Fragment

Switch to the SV40 Display window and activate its Graphics Pane. Open the Fragment Wizard again:

- **First screen:** Select the Design Donor Fragment option on the first screen; click the Next button.
- **Second screen:** Click on the LARGE_T signal’s symbol or label in the Graphics Pane to select it. Click the Finish button in the Fragment Wizard.

Inspect the donor fragment in the New Fragment message box, and click Add to List button. (Figure 11.2) The donor fragment is added to the Molecule Goal List.

4. Inspect the Goal Molecule Definition List

Click the Open Goal List button ( ) in the main toolbar, opening the Lists dialog box. The Fragments section of the dialog box contains the Goal Molecule Definition List, listing the two fragments you defined. The recipient fragment (BPVI) must be the first in the Goal Molecule Definition List. (If it is not, select the fragment, then click the Up button to move it into the correct position.)
Highlight the SV40 fragment and click the **Edit** button, opening the Fragment Editor dialog box. Click the **Inverted** check box to change LARGE_T's direction to match the recipient's direction and click **OK**.

You could leave LARGE_T in its original orientation if you want to; the system will design your new molecule either way. We have changed LARGE_T to Inverted only to demonstrate that Vector NTI can clone fragments in different orientations.

### 5. Enter General Information for Your New Molecule

Click the **Run** button. Enter the name and description for your new molecule as you learned in previous tutorials:

- Name the molecule **TUTORIAL 3**.
- Click the **General Info** button, opening the General data dialog box.
- In the Description field, enter **Tutorial molecule #3**.
- Set the Replicon Type to Plasmid.
- Turn on the Bacteria Extra-Chromosome Replication option.
- Enter your name as a keyword.
- Click **OK** to return to the Design Molecule dialog box.
- Check the **Recipient's Start** button in the radio button group to position the new molecule start at the same place (if possible) where the recipient molecule (BPV1) starts.

### 6. Prepare to Design the New Molecule

Click the **Design** button in the upper right corner of the Design Molecule dialog box. When asked for a subset name, select the **Tutorial** subset you created before; click **OK** to continue.

The Design Parameters dialog box appears. Leave all the settings at their default values and move on to the next step.

### 7. Set the Design Preferences

Click on the **Preferences** button, opening the Design Preferences dialog box. Note that the blunt-blunt ligation box is already turned off. Vector NTI remembers your previous design preferences so that you do not have to set them every time you design a new molecule.

Below the check boxes are priority lists specifying which of the techniques are preferable. Leave these at their default values.

Click the **OK** button to accept the Design Preferences and return to the Design Parameters dialog box.

### 8. Design the New Molecule

Click the **Start Design** button. As before, an optimum cloning method is quickly found, and the goal molecule is constructed based on that best option.
9. Inspect the New Molecule

When the design is complete Vector NTI opens a new Molecule Display window containing the molecule you have created (Figure 11.2).

![Figure 11.2 Inspecting the new molecule you have created](image)

10. Inspect the Graphical Map and Text Description of Your New Molecule

To verify the restriction enzymes used in the design process, they must be specified in the restriction enzyme display list. This is accomplished by clicking the Display Setup button on the Pane Selection toolbar.

In the Molecule Display Setup dialog box, check the Restriction Map checkbox and press the RMap Setup button to specify the restriction sites to be displayed. The Restriction Map Setup dialog box displays in the Use Enzymes column the restriction enzymes, whose sites are shown by default. To include the enzymes from the design process in the display list, click the < Add button.

In the Choose Database Enzymes dialog box, with the Palindromes/Non-ambiguous subset selected in the Look in: pull-down menu at the top of the dialog box, select AccEBI and M.NgoAV from the list of available restriction enzymes. Click OK. For more information, See Restriction Map Setup on page 219.

Activate the Text Pane, and open TUTORIAL 3’s Design Description folder and the Step #1 sub-folder. Review the description of how to create TUTORIAL 3. In particular, note the following:

- You defined a specific ApaLI site on the recipient molecule as the site into which the donor fragment was to be cloned. You also requested that Vector NTI save this site in the recombinant molecule. ApaLI was not a unique site, so Vector NTI had to use partial digestion to isolate the recipient. Then the system had to find a way to insert the donor fragment. There may not have been any good sites available on the donor fragment to fit into the recipient’s ApaLI site; therefore, Vector NTI excised the donor fragment with convenient sites that did not require partial digestion and left short flank regions, AccEBI and M.NgoAV. These termini were then filled in completely and ApaLI linkers were attached to the blunt ends. This fragment was then cloned into the recipient.
You can be sure the cloned fragment contains no ApaLI sites (Vector NTI checked this when selecting a linker) and that there were no easier ways to perform this cloning.

The rest of TUTORIAL 3’s Design Description is similar to the Design Description of TUTORIAL 2, designed in the previous tutorial. However, because the 5’ and 3’ termini of the recipient for TUTORIAL 3 are identical, we need to know how to distinguish clones with the proper orientation. Thus a new paragraph appears in TUTORIAL 3’s design description, recommending that the AauI and ApaLI restriction sites would allow you to distinguish between different insert orientations upon gel analysis.

You may print out this design plan if you wish, and then proceed to the next design task.

Second Design: Complex Donor Fragment

You will now design a second molecule based on BPV1 and SV40, this time making the donor more complicated rather than the recipient.

1. Define Recipient

Return to the BPV1 Display window’s Graphics Pane and open the Fragment Wizard:

- **First screen**: Select the Design Recipient Fragment option and click the Next button.
- **Second screen**: Select the Set to a Position option and enter 5000 as the position of the 5’ terminus. Click the Next button.
- **Third screen**: To define the 3’ terminus, select the Set to a Position option again; enter 2500. Click the Finish button.

Inspect the results in the message box, and click the Add to List button. The recipient fragment is added to the Molecule Goal List.

2. Define Donor

Switch to the SV40 Display window’s Graphics Pane and open the Fragment Wizard.

- **First screen**: Select the Design Donor Fragment option and click the Next button.
- **Second screen**: Click on the LARGE_T signal’s symbol or label in the Graphics Pane to select it. Click the Next button.
- **Third screen**: The default option on the next screen is “Leave terminus Undefined.” Do not change this option; click the Next button.
- **Fourth screen**: On this flank region screen, you can specify the maximum size of the flank region or let Vector NTI use all available space outside the selected signal. Select the **Use flank region no larger than** option. You can either enter the maximum length in the text box or select the fragment in the Graphics Pane. Move your cursor over the 5’ end of the selection wire-frame in the Graphics Pane (indicated by the symbol). Click and drag the 5’ end to the position at about nucleotide 2250. The current position is displayed in the status bar. (Remember you can use SHIFT + right and left arrow keys to fine-tune the selection.) The edit box in the Fragment Wizard shows you the maximum size of the flank region (it should be about 400 nucleotides). When you are finished, click the Next button to go to the 3’ terminus screen.
- **Fifth screen**: Specify here that Vector NTI should use the Ncol restriction site at nucleotide 38 to cut the 3’ terminus of the donor fragment. Select the **Use specific site** option, SHIFT + CLICK on the Ncol restriction site at nucleotide 38. (Drag the Fragment Wizard out of the way, if necessary). The name and the position of the restriction site appear in the
Fragment wizard screen. Click the **Finish** button to complete the definition of the donor fragment.

In the New Fragment message box, inspect the selected donor fragment. Note that the 5’ terminus is described as “Undefined,” with the specified maximum length of a flank region and the 3’ terminus is set to the _NcoI_ site (Figure 11.3). Click the **Add to List** button, adding the donor fragment to the Molecule Goal list.

3. **Inspect the Goal List**

Click the **Open Goal List** button to open the List dialog box and inspect the fragments you have defined. The recipient fragment (BPV1) must be listed first in the Goal Molecule Definition List. Select the donor fragment (SV40) and press the **Edit** button, opening the Fragment Editor. One end of the donor is determined by a restriction site, while the other end has a flank region defined. This makes the situation considerably more complicated than our previous cases, where the donors were more simply defined. Click the **Inverted** check box and click **OK**.

4. **Design the New Molecule**

Click the **Run** button. Enter the name **TUTORIAL 4** and enter the other necessary information about the new molecule. Check the **Recipient’s Start** button to make the start of the new molecule at the same place (if possible) where the recipient molecule (BPV1) starts.

Click the **Design** button in the upper right corner of the dialog box. Select the “Tutorial” subset for the molecule and click **OK** to continue. Leave the Design Preferences at their default settings and click the **Start Design** button. After a few moments, Vector NTI produces a workable design for the molecule you have requested.

5. **Inspect and Print the New Molecule**

To verify the restriction enzymes used in the design process, as mentioned in the previous tutorial, they must be specified in the restriction enzyme display list. Using the skills you learned previously in this tutorial, add AccEBl to the restriction enzyme display list.

In the new Molecule Display window containing TUTORIAL 4, inspect the Step #1 subfolder of the Design Description folder. In this case, despite the more complex conditions you specified, Vector NTI found a simple and convenient design. Both donor and recipient can be isolated at the NcoI and AccEBl sites, so the cloning is straightforward.

You specified a maximum flank region of about 400 nucleotides; Vector NTI found a good AccEBl site, allowing the actual flank region to stay within the assigned limit.

If you wish, print out the design description and/or graphical map of TUTORIAL 4.
Introduction

This chapter introduces you to Vector NTI’s Gel Display windows, powerful tools for running and analyzing Gel electrophoresis. Vector NTI predicts the results of actual Gel electrophoresis experiments and displays them in text and graphics form.

At the end of this tutorial session, you will be able to:

- Open a new Gel Display window
- Create a gel marker (size standard)
- Add samples and gel markers to a “virtual” gel
- “Run” the gel and animate it
- Calculate the time needed to separate a specific fragment
- Manipulate gel graphics

Follow the steps in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch Vector NTI

Launch Vector NTI using techniques learned in previous tutorials. Open the Vector NTI workspace window and maximize it.

2. Create a New Gel and Open a Display Window

Click the New Gel button ( ) on the Main Toolbar to create a new Gel. In the Gel Setup dialog box that opens (Figure 12.1), you can define your own gel and electrophoresis parameters, such as:
- gel type (agarose, polyacrylamide, ...)
- gel concentration
- gel dimensions
- voltage
- electric field type (steady, pulse)

![Figure 12.1 Gel Setup dialog box, where you can define your own gel and electrophoresis parameters](image)

Leave all of the default gel settings.

From the Electrophoresis Profile drop down menu, select the **Examples of Agarose Gel** and click **OK**. Your new gel opens in the Gel Display window. Maximize the window.

### 3. Create a Sample and Add it to the Gel

The Display window that opens displays an empty gel ready to receive samples. Your sample will consist of the SV40 molecule cut by the HaeIII and HindIII restriction endonucleases.

Click the **Create Sample** button on the Window toolbar. In the Create Gel Sample dialog box that opens (Figure 12.2), in the Molecules box, select SV40. In the Enzymes box, scroll to select HaeIII and HindIII. As you select the molecule and RENs, their names appear in the Selected for Gel Sample column on the right side of the dialog box.
In the Sample Name box near the bottom of the dialog box, enter **Sample 1**. In the Description box, enter **SV40 cut by Haelll and Hindill**.

![Figure 12.2 Create Gel Sample dialog box](image)

Click the **Add to Gel** button in the upper left corner of the dialog box. Vector NTI cuts SV40 with the two RENs and adds the resulting fragments to the gel shown in the active Gel Display window. Click the **Close** button to close the dialog box.

4. **Add Gel Markers**

Now let’s add three gel markers as size standards for the gel.

Click the **Add Marker Lane** button ( ) on the Window toolbar. The dialog box that opens displays the gel markers currently stored in the database. Add the following gel markers:

- **Lane 2**: Select the PBR322-Haelll gel marker; click **OK**.
- **Lane 3**: Select SPP-EcoRI; click **OK**.

No lambda-HindIII size marker appears in the Vector NTI database. Because it’s a commonly used size standard, let’s create our own. Select **Gel > Create Gel Marker** on the menu bar. In the New Gel Marker dialog box, on the General tab, enter Lambda HindIII. On the Gel Marker tab, enter each fragment size, then click the **Add** button: 23130, 9416, 6557, 4361, 2322, 2027, 560. Click **OK**.

- **Lane 4**: Select Lambda HindIII, your newly created gel marker. Click **OK**.
5. Arrange the Gel Display Window Conveniently

Using the split bars, scroll bars and Zoom buttons, arrange the Gel Display window to your liking (Figure 12.3).

![Figure 12.3 Arranging the Gel Display window](image)

6. Inspect the Gel Display Window Text Pane

Activate the Text Pane of your Gel Display window by clicking the Text Pane button ( ) and inspect its contents. The General Description folder describes the electrophoresis parameters. There is also a folder for each lane of the gel.

Open the folder “1. Sample 1” by double clicking on it or clicking on [+]. Open the folder representing the longest fragment in the sample, labeled 861. For this fragment, the source molecule is SV40, and the ends of the fragment are the HindIII site found at position 4003 and the HaeIII site found at position 4864. Note that the source of the fragment is described. There are also subfolders for each fragment in decreasing order of length. Close the Sample 1 folder with a double-click.

Double click the folder labeled “SPPI – EcoRI” the gel marker found in lane 3 of the gel, and open the subfolder “8000”. Note that it is empty. The folder is empty because for gel marker fragments, only the length is stored in the database.

7. Change Display Color for a Fragment

You can use the Text Pane to change the color and style for displaying fragments. Open the SAMPLE1 folder again. Scroll down to a folder labeled “540 [length of fragment]” and open its associated shortcut menu.

Choose Sample Fragment Properties. In the dialog box that opens, you can choose line style and color for this fragment. Leave the line unchanged, but change the color to green.

As the gel runs, you can locate the 540-bp fragment by looking for the green color.
To run the gel, activate the gel pane. Click the Fit To Window button ( ), causing the gel display to fit exactly within the current window.

Click the Step Forward button ( ) in the Window Toolbar. The time indicator increases by fifteen minutes, and the fragments advance in the "gel". Click the Step Forward button several more times.

Click the Step Back button ( ) in the Window toolbar. The time indicator and gel display now back up.

Now click once in the time indicator box to activate it. Enter 1:30 in the time indicator box, and press Enter. The gel display is set for 1 hour and 30 minutes of elapsed time.

As the gel advances and retreats, note how the scale at the left of the gel keeps track of the fragment sizes (Figure 12.4).

Click the Animate button ( ) on the window toolbar to view a continuous simulation of a Gel electrophoresis. The gel now runs continuously at the speed specified in the Gel Setup dialog box when you created the gel.

When the gel time indicator reaches three or four hours, click the Animate button again to stop the run. The fragments are now well separated.

Click the Zoom In button ( ) several times until fuzzy gray bars appear around the lines representing the DNA fragments (Figure 12.5). Using the scroll bar, move down along the gel until
your green fragment appears. The gray bars show the minimum distance by which fragments must be separated so that you can distinguish them in the gel.

Note that the gray bars of many fragments touch each other. This indicates that you will not be able to distinguish those fragments in the gel. The size of the gray bars is determined by the Separation Distance parameter, defined in the Gel And Display Setup dialog box described in step 2.

![Figure 12.5 Optimizing the Gel Display](image)

Click the **True-Scale View** button ( ), which sizes the gel to represent its true scale in a laboratory.

Click the **Fit To Window** button ( ) again, and the graphical gel display is sized to fit exactly within the current window.

### 11. Change to a Polyacrylamide Gel from the Electrophoresis Profile

Change the parameters of the gel by clicking on the **Gel and Display Setup** button ( ) in the Window Toolbar. In the Electrophoresis Profile drop-down menu in the dialog box, select **Example of Polyacrylamide Gel** and click **OK**. The electrophoresis parameters are changed. While the contents of the lane have not changed, as you view your gel again, note that the positions of the various fragments have been altered.

### 12. Estimate Fragment Separation Time

Some fragments in our gel are still too close together to distinguish. To estimate how long the gel would have to run to separate them, use Vector NTI's built in calculator as follows:

- Activate the Graphics Pane.
- In the Time Indicator box of the Window Toolbar, type in 2 hours and 30 minutes (2:30).
- Click ENTER.

Click the **Fit to Window** button ( ) to set the gel to its standard size and press the **Zoom In** button ( ) three times to magnify the view of the gel.
Review the first lane of the gel, containing SAMPLE1. With the scroll bar, scroll down until you find the section of the gel containing fragments with lengths between 500 and 700 bp. (You will see three fragments just above the 500-bp mark.) The middle fragment (length 540 bp) is green, because you set its display color in step 7 of this tutorial. The fragments’ gray borders overlap with each other, indicating that these fragments are still too close together to be distinguishable on the real gel.

Move the cursor to Lane 1, where it changes into a horizontal I-beam ( ). Place the mouse cursor above the three fragments. Left-click + drag the mouse (I-beam) downwards, highlighting all three unseparated bands.

Click the Calculate Separation Time button ( ) in the Window Toolbar. A message appears, telling you that the gel would have to run for 5 hours and 30 minutes to separate the selected fragments. This gives you a general idea of what to expect in an actual gel run. Click OK. The gel moves immediately to the position at that time elapse. Scroll down so you can view the fragments.

Click anywhere in the gel pane to remove the selection.

13. Save the Gel Display Window to a Gel Document File

To save the data in the Gel Display window, including gel settings and fragments positions, select Gel > Save as Document from the menu bar.

A dialog box appears, asking you for a filename. Name the file Tutorial.gd and click the Save button. The gel is saved in a Gel Document file.

To close the Gel window, select Gel > Close.

Now open the saved Gel Document you created. Select Gel > Open Document, and select the name of the saved document file from the file list. Your Gel Document opens, and all your fragments and settings have been preserved.

14. Close the Gel Display Window

You have now taken a first look at Vector NTI’s Gel Display windows and restriction digestion features. Close the window using Gel > Close.

15. Calculate Restriction Fragments

One of the simple options in Vector NTI is creating restriction fragments of molecules.

Using techniques you have learned, open and maximize the pATH23 Molecule Display window. Select Analyses > Restriction Analyses > Restriction Fragments on the menu bar. Deselect all of the RENs listed. Then select ApaLI and EcoRI. Click OK.

Inspecting the Text Pane reveals that all of your resulting fragments are noted, with their size, source, and the restriction sites producing those fragments.

Identify several of the restriction fragments on the molecule map by placing the cursor on the appropriate line in the Restriction Fragments folder (Text Pane) and opening the shortcut menu. Select Find Fragment. Clicking this option results in the selection of the fragment on the graphics map. Close the windows.
Introduction

This chapter introduces you to Vector NTI tools that provide connections to external programs and Internet services.

At the end of this tutorial session, you will be able to:

- Send data from Vector NTI to several public Internet servers
- Perform a simple BLAST search, alignment and analysis
- Save Internet search results as Comments in a database file

Follow the steps in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch Vector NTI and Open pBR322 in a New Display Window

Launch Vector NTI by selecting Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI. Click the Open Document button ( ) on the main toolbar. Open the Molecule Viewer for pBR322 using the techniques you learned in previous tutorials.

2. Select the Whole Sequence of pBR322 and Use the BLAST Search Tool

Select Analyses > Web Analyses > Compare Against > Many databases using BLAST @ NCBI.

Alternatively, BLAST Searches can be launched in Analysis Monitor (page 569), or by selecting Start > Programs > Invitrogen > Vector NTI Suite 10 > BLAST Search.

Note: Some World Wide Web browsers require the Internet connection to be established before you start the browser. If your TCP/IP stack or other Internet connection software cannot be started "on the fly," you need to connect before selecting any of Vector NTI's Internet-related commands.
In the Sequence Data dialog box of the BLAST viewer (Figure 13.1), choose the range and strand of the sequence to be sent to the server: Whole Sequence/Direct Strand.

![Sequence Data dialog box of the BLAST viewer](image)

Figure 13.1 Sequence Data dialog box of the BLAST viewer

Click OK to continue. Your selected browser may open automatically or a list of available browsers may open.

If appropriate, select the name of your Internet browser from the list and click OK to continue. This configures Vector NTI's connection to the Internet for all subsequent connections.

**Note:** The standard Vector NTI distribution contains support modules for many popular Internet browsers. If your Internet browser is not in the list, try selecting one of the “Autodetect” modules from the displayed list. The latest additions to the list are available on the Vector NTI web page: [www.invitrogen.com/bioinformatics](http://www.invitrogen.com/bioinformatics)

Vector NTI opens the BLAST Search page in the Internet browser (Figure 13.2). (The page may look only similar to this, depending on your browser).

![NCBI Blast Search page](image)

Figure 13.2 NCBI Blast Search page

This page is Vector NTI's gateway to the NCBI BLAST Search server and contains options that can be used to change various BLAST search parameters.

The sequence of pBR322 is already entered in the Search field. In the Choose database section, select the nr database. Scroll to the bottom of the page. If your e-mail address appears in the Send results by e-mail text box, delete it. Click the BLAST! button.

**Note:** At this step, your Internet browser may display a dialog box or start a program to establish an Internet connection. If this happens, enter the required parameters and continue.

Once an Internet connection with the NCBI server has been established, you should see the NCBI formatting BLAST page. To proceed, click the Format! button on that page. The response time for the BLAST results varies depending on the server load. If the server is busy, wait several minutes and resubmit the query.
Note: Since the interface of Internet servers changes from time to time, you may receive a message that the requested resource is not found. This means that Vector NTI’s server gateway page is obsolete.

When you receive the results from the BLAST server, click on the second molecule link (at the time of this writing (9/2003), the first hit was pBR322 itself and second hit was U03501/YRP7). Another page appears, containing the GenBank description of the molecule (Figure 13.3):

![Figure 13.3 NCBI Web page containing the GenBank description of the molecule](image)

3. Display the Result of the Query in a Molecule Display Window

To transfer the result of the query back to Vector NTI, highlight the GenBank text starting at the word “LOCUS” and ending with the “//” at the end of the sequence. Select **Edit > Copy** from the drop-down menu (or its analog) in your browser to copy the selected text to the Clipboard. Switch back to the Vector NTI workspace and choose **Edit > Open from Clipboard** from the Vector NTI Main menu (Figure 13.4).

![Figure 13.4 Transferring the result of the query back to Vector NTI](image)
In the Open/Import from ‘Clipboard’ dialog box that opens, select **GenBank file** from the list of file types and click the **Open** button (Figure 13.5).

![Open/Import from "Clipboard" dialog box](Image)

**Figure 13.5 Selecting the file type for the transferred query**

Vector NTI opens your molecule in a new Molecule Viewing window, automatically generating the restriction map and graphics representation of the molecule.

4. **Save (Send to) or Open the Result of the Query Using the Save Button**

An alternative way to transfer search results from the NCBI server to Vector NTI is to use the **Send to** button on the query’s GenBank report. You can also use the **Send to** button to save the query molecule to your hard drive as an independent file.

Switch back to the U03501/YRP7 GenBank report and locate the **Send to** button just above the text of the GenBank report. Click the **Send to** button. In the drop-down menu that appears alongside the **Send to** button, make sure the **File** option is selected. The File Download dialog box appears.

- If you choose the **Open** option, your browser will try to load the document into the associated application. The Open With dialog box appears, scroll down the list of available programs and select Vector NTI. Click **OK**. In the Open/Import dialog box, select **GenBank file** from the list of file types and click the **Open** button (Figure 13.5). If Vector NTI was configured correctly, a new window should appear in its workspace showing the U03501 molecule.

- If you choose the **Save** option, you can specify a file name and a location to store the U03501 molecule as a file on your hard drive.

5. **Use Alignment and Analysis Tools**

The standard distribution of Vector NTI has several tools to perform sequence alignment and analysis. To perform the alignment, switch to the Database Explorer window using the **Local Database** button ( ). Select the Proteins table from the drop-down menu.

Vector NTI displays the list of proteins in the database **CTRL + CLICK** on 41BB_HUMAN and 41BB_MOUSE to select both. Choose **Align > Web Alignment Tools > Multiple Alignments > Multiple Alignment @ BCM Server** from the Database Explorer menu bar. Vector NTI opens
the browser and displays the BCM Search page containing the sequences of the selected proteins in FASTA format (Figure 13.6).

Click the Submit button to start the ClustalW alignment (the default option). If the BCM server is not too busy, you will receive the response in 5-10 seconds. Scroll through the screen to review the text-based alignment results.

To analyze a protein using the PROSITE program on the ExPASy server in Switzerland, return to the Database Explorer window and click on the 41BB_HUMAN protein again to select it. Choose Analyses > Web Analyses > Protein Feature Search > Domain Database PROSITE using ScanProsite @ ExPASy from the Database Explorer’s menu bar.

When you see the ScanProsite page containing the sequence of 41BB_HUMAN in the browser, select and copy the sequence from the Scan a Protein for PROSITE matches text entry box by right-clicking and opening a short-cut menu. From the task bar of the ScanProsite webpage, select PROSITE. In the PROSITE: Database of Protein Families and Domains, scroll down to the Tools for PROSITE section and paste the sequence into the text entry box by right-clicking the mouse. Click the Quick Scan button. If the ExPASy server is not too busy, you will receive the response in 2-5 seconds.

Searches are performed in PROSITE for biologically significant protein patterns and profiles for known protein families and domains. When the search is complete, the results are displayed on the ScanProsite webpage. Scroll the page to review the pattern matches listed in their numerical...
and graphical form. For details, click on the site ID links, such as PDOC00001 for the glycosylation site, opened and illustrated in Figure 13.7.

![Figure 13.7 Site ID link for PDOC00001 (glycosylation site)](image)

### 6. Saving Search Results as Comments

Vector NTI allows you to associate any text data with database objects in the form of comments. Let's save the search results to the database for future reference.

Select the resulting hits, starting with hit [1], by dragging the mouse across and down the page. Do not include the sequence, which is already stored in the Vector NTI database.

Choose **Edit > Copy** or press **CTRL-C** to copy results to the Clipboard Return to Database Explorer, and with 41_BB HUMAN still selected in the Database Explorer window, select **Protein > Edit** from the Explorer’s menu bar.

In the tabbed Edit dialog box that opens, you can edit various information associated with the protein molecule (Figure 13.8). Click on the **Comments** tab, scroll to a point at the end of the existing comment text and click to set the text cursor for insertion.

![Figure 13.8 Edit dialog box, where you can edit information associated with the protein molecule](image)
Now press CTRL-V to paste the search results from the Clipboard.

Edit the resulting text to your liking and click the OK button to save it to the database (Figure 13.9).

The search results are now saved in the database for later retrieval. To view the comments from 41_HUMAN's Display window, double-click on the Comments line in the Text Pane of a Molecule Display window for 41_Human.

7. Close the Molecule Display Windows and Exit Vector NTI

You have now taken a first look at Vector NTI's Tools for Internet connection. To close all display windows, select File > Close; to exit the program, select File > Exit from the menu bar. If you are not going to continue your work with the Internet, close the Internet browser and disconnect, if necessary.
Introduction

This chapter introduces you to operations of AlignX. The simultaneous alignment of many nucleotide or amino acid sequences is an essential tool in molecular biology. Multiple alignments are used to find diagnostic patterns, characterize protein families, as well as to detect or demonstrate a similarity between new sequences and existing families of sequences. They are also useful in predicting secondary and tertiary structures of new sequences, suggesting oligonucleotide primers for PCR and serving as an essential prelude to molecular evolutionary analysis.

At the end of this tutorial session, you will be able to:

- Import molecules from other Vector NTI Advance applications into AlignX
- Describe elements of the AlignX user interface
- Create an alignment, and add molecules to an alignment
- Modify alignment graphics and sequence
- Edit an alignment
- Create and interpret an AlignX Dot Matrix
- Export molecules from AlignX into other Vector NTI applications

Follow the steps of the tutorial in the order shown. Figures show what your screen should look like at various points along the way.

1. Launch AlignX from the Database Explorer

Using techniques you learned in previous tutorials, open the Vector NTI Database Explorer. Now open AlignX from the Start menu by selecting Start > Program > Invitrogen > Vector NTI Suite 10 > AlignX. Alternatively, you can open AlignX from the Database Explorer by selecting Align > AlignX - Open New Alignment Window. In the Database Explorer, activate the Protein table by selecting Table > Protein Molecules and select the Protein Molecules (MAIN) subset. In the Database Objects List Pane, scroll to the molecules 5HIA for human, rat, and mouse.
Click the cursor to the right of the bottom of the last of the three proteins, and drag a box around all three molecules. This is a new way to select a group of objects in a subset list (Figure 14.1).

Now position the Database Explorer window and the AlignX window side by side, and click and drag the selected molecules into the AlignX viewer.

Next, in the Database Explorer protein list, select the 41_BB Human molecule and drag it into AlignX.

An alternative method is to select the molecules in Database Explorer and select Align > AlignX – Align Selected Molecules.

2. Examine AlignX Display Window

When you open AlignX, even after bringing molecules in, three of the panes are empty. After you initiate the alignment, the alignment results will appear in the empty panes.

In the Text Pane, select the 5H1A molecules. Click the Align button or choose Align > Align Selected Sequences from the drop-down menu, initiating the alignment that now appears in the display window.
The AlignX Display window consists of a menu bar, three toolbars and context menus specific for each pane (Figure 14.2). The AlignX Display window includes the Text Pane, the Guide Tree Pane (still empty), the Analysis Pane and the Alignment Pane. The panes are divided by split bars that can be moved to resize the panes.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

As in the other Vector NTI Advance applications, to apply a command from the main menu in a given pane (Print, Print Preview, Camera), that pane must be active. The active pane can be toggled with the **Switch Panes** buttons in the Pane Selection Toolbar or by clicking anywhere in the pane you want to activate.

In the Text Pane, double click on a molecule folder to open and review it. Move the split bar at the bottom of the pane and open some of the subfolders. A Text Pane folder and its subfolders contain a thorough description of the molecule. Data includes molecule type and form, user-defined fields (from Vector NTI) and standard fields (such as GenBank), comments, references and so forth.

The default Analysis Pane contains three graphical representations of alignment results. Move the lower split bar and the split bars between the graphs and/or use the scroll bar to better view them.

- The first graph displays the alignment quality profile (Similarity). The specific values (in a 0-1 range) are assigned to each residue at a given alignment position in each aligned sequenced depending on whether the residue is identical, similar or weakly similar to the corresponding residue of the consensus sequence. The values for each residue at a given position are added together and the number of the sequences in the alignment normalizes the resulting value. The default values are 1, 0.5 and 0.2 for identical, similar and weakly similar residues respectively.

- The second graph displays the statistical significance profile (Absolute Complexity) of an alignment. It is calculated as a sum of all pairwise residue substitution scores at a given alignment position normalized by the number of pairs in the alignment. The scores are taken from the residue substitution matrix used for alignment calculation.

- The third graph displays the statistical significance (Absolute Complexity) of an alignment for a selected molecule to the consensus sequence. The graph is calculated in a manner identical to the one in the second graph.

The Alignment Pane is described in step 4.

You can add additional analyses to the graphics pane by activating the Graphics Pane and choosing **View > List of Analyses** from the drop down menu to open the Analysis List setup dialog box (Figure 14.3):

![Figure 14.3 Protein Alignment Analysis List setup dialog box](image)
The dialog box lists all of the analyses available for DNA or Protein, depending on the type of aligned molecules. Currently displayed analyses are listed in the Selected Analyses box. To add an analysis, select it in the Available Analyses list box and click the **Add Analysis** button ( ) to move it to the Selected Analyses list. Alternately, you can move an analysis to the lower list by double clicking on it.

The plot display order is determined by the order the analyses are listed in the Analysis List setup box. To modify the order, select an analysis and move it up by clicking the **Move Up** button ( ). Click **OK** to close the Analysis list setup dialog box and generate the new analyses.

### 3. Add Molecule to Alignment

To add a molecule to your alignment, select the 41_BB Human molecule in the Text Pane and click the **Add to Alignment** button ( ). You can add as many molecules as you like to the alignment as long as they are listed in the Text Pane.

### 4. Inspect the Guide Tree

Note that now the Guide Tree Pane shows in its specific pane (Figure 14.4). Phylogenetic analysis is the means of studying presumed evolutionary relationships. An inferred evolutionary history is displayed in a treelike diagram suggesting the inherited relationships between the molecules. In Vector NTI, a guide tree is only calculated when there are more than 3 molecules in the alignment.

![Figure 14.4 Guide Tree Pane](image)

The guide tree, which resembles a phylogenetic tree, is built using the Neighbor Joining method (NJ) of Saitou and Nei. The NJ method works on a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between the sequences. The guide tree is calculated after the sequences are aligned. AlignX displays the calculated distance values in parenthesis following the molecule name.

### 5. Review the Alignment Parameters and Matrix

Click the **Alignment Setup** button ( ) on the Viewer Toolbar to see how you can adjust different parameters for the alignment. Select each of the tabs in the dialog box to quickly review the settings. Leave the last tab, Score Matrix selected.

All algorithms designed to evaluate pairwise sequence alignment are based on systems that assign scores to aligned residues, detecting similarities between differing sequences. This page displays the matrix, blosum 62mt2, used to generate your multiple alignment. Amino acids that are identical or similar in alignment score higher than those less similar. The most common of many different scoring systems are based on substitutions of amino acids in related proteins.
BLOSUM matrices are calculated from local, ungapped alignments of distantly related sequences. The matrix scores represent the minimum percent identity of the blocks used to construct the matrix; greater numbers represent lesser distances.

In Vector NTI, you can modify matrices using the Matrix Editor. Customized matrices can also be devised and stored in the database for shared or private use.

Leave the settings unchanged. It is a good idea when performing the various analyses in Vector NTI applications to run them first using the default settings. Review the results, then change the parameters as you like before running the analyses again.

6. Examine the Alignment Pane

Now move the split bars up to optimizing viewing of the Alignment Pane. Sequences of all the molecules included in the alignment are displayed here, with their names listed at the left of the corresponding sequence. Residues in the alignment are colored according to the following scheme:

<table>
<thead>
<tr>
<th>Alignment Color</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>black on window default color</td>
<td>non-similar residues</td>
</tr>
<tr>
<td>blue on cyan</td>
<td>consensus residue derived from a block of similar residues at a given position</td>
</tr>
<tr>
<td>black on green</td>
<td>consensus residue derived from the occurrence of greater than 50% of a single residue at a given position</td>
</tr>
<tr>
<td>red on yellow</td>
<td>consensus residue derived from a completely conserved residue at a given position</td>
</tr>
<tr>
<td>green on window default color</td>
<td>residue weakly similar to consensus residue at given position</td>
</tr>
</tbody>
</table>

Table 14.1 Alignment color scheme

Move the horizontal scroll bar at the bottom of the Alignment Pane. Drag the mouse across some of the sequence, selecting it. Note that the corresponding region is selected in the Analysis Pane as you do so.

The sequence names remain stationary at the left of the Alignment Pane no matter how the sequences are repositioned horizontally. If the list of molecules listed in the Alignment Pane is longer than the size of the pane, a vertical scroll bar appears as well. Vertical scrolling affects both sequences and their names but the consensus sequence remains visible at the bottom of the pane.
7. Edit the Alignment

After inspecting the alignment, click the **Edit Alignment** button ( ) on the Viewer Toolbar to open the Alignment Editor dialog box (Figure 14.5). This dialog box displays rows of selected molecule names and sequences.

![Figure 14.5 Alignment Editor dialog box, which displays rows of selected molecule names and sequences](image)

Scroll the sequences to find the 3' end of the 41_BB Human molecule. Click and drag to select **EEEGGCEL**. Note that as you do so, two of the Move Selected Block buttons in the bottom of the dialog box become enabled. The **Move Selected Block** text box indicates the number of positions at a time to move, and defaults to one. Click the ( ) button, which moves the selected block one space to the right, as indicated in the Move Selected Block text box. (The ( ) button would move the block to the extreme right of the gap.) *Because editing an alignment can only occur when selected residue(s) are shifted into an adjacent gap, the arrows are only enabled when the selection block is next to a gap into which the selection may be shifted.*

8. Generate Dot Matrix Plot

To perform a Dot Matrix analysis, select open a Dot Matrix window by clicking the Dot Matrix ( ) button on the main toolbar. A Dot Matrix window opens in a separate modal popup window (Figure 14.6). In the drop-down menu at the top of the screen, select 5H1A HUMAN (the first selected molecule in the AlignX window.) In the second drop-down menu, select, 5H1A mouse, launching the Dot Matrix.

![Figure 14.6 Dot Matrix analysis](image)
The Dot Matrix analysis is primarily a method for finding all possible matches of residues between two sequences. One sequence (A) is listed across the top of a page and the other sequence (B) is listed down the left side.

The graphical representation of a Dot Matrix is a dashed rectangle in which any region of similar sequence is revealed by a diagonal row of dots. The molecule selected first corresponds to the horizontal axis, its name being shown in the right top corner of the Dot Matrix rectangle. The second molecule corresponds to the vertical axis. Each axis has its own scale that initially displays the sequence positions.

When the Dot Matrix is first calculated, it is shown as a representation of the full length of a molecule. Because it is difficult to analyze specific molecule regions when several thousand residues are represented, click and hold the left mouse button within the Dot Matrix rectangle while dragging the mouse, displaying the tracking dotted rectangle (Figure 14.7).

![Figure 14.7 Selected region of the Dot Matrix](image1)

When the mouse button is released, the selected region is enlarged to the entire Dot Matrix rectangle (Figure 14.8).

![Figure 14.8 Enlarged region of the Dot Matrix](image2)

Click Undo Zoom ( ) to return to the previous screen. Close the Dot Matrix window.
9. Export Alignment Molecules

AlignX is fully integrated with other components of Vector NTI Advance. As an example of exporting a molecule from AlignX into other applications, select the molecule 5H1A_HUMAN in the Text Pane and select **Analyses > Back Translation**. In the Sequence Data dialog box, enter the Range (Whole Sequence or Selection Only). Click **OK**. The display window that opens displays the original amino acid sequence in the upper pane, and the back translated sequence in the lower pane. In Figure 14.9, that sequence displayed is the “most ambiguous” as shown by the scale above the sequence.

![Figure 14.9 Back translated sequence](image)

Drag the scale button to the right, noting how the translation code is modified as you proceed. Uncheck the “3 letters in translation” box at the bottom, changing the amino acid code to 1-letter representation. Exit the Back Translation mode. (For more information on the Back Translation function, see page 297.)

Now select **Analyses > BioAnnotator > Analyze Selected Molecule**. The BioAnnotator display window opens, displaying physiochemical characteristics of the 5H1A_HUMAN protein. BioAnnotator will be discussed in the next tutorial. Close the BioAnnotator window.

10. Copy and Print an AlignX Pane

The Copy and Print features work the same in AlignX as in the other Vector NTI components. To copy the contents of a pane and paste them exactly as they appear in AlignX, activate the pane and click the **Camera** button ( ). Paste the copied material in any other application, such as a word processing document.

Print pane contents in the usual manner, by pressing the **Print** button ( ).

11. Save the Project and Close AlignX

You can save your project if you choose by selecting **Project > Save As** and selecting your project destination in the Save As dialog box. The next time you open AlignX, the last several opened projects are listed under the Project menu option. Selecting a project from that list opens it.
Close AlignX in the same way you close other applications of Vector NTI: select **Project > Exit**. This ends your tutorial session with AlignX.
CHAPTER 15

TUTORIAL: BIOANNOTATOR

Introduction

This chapter introduces you to the Vector NTI Advance application, BioAnnotator. BioAnnotator performs certain types of DNA and protein sequence analyses and displays the results as linear graphics. This allows you to visualize the physiochemical characteristics of a molecule, which may be useful for inferring the potential structure/function of molecule regions.

At the end of this tutorial session, you will be able to:

- Launch BioAnnotator for a molecule you create by translating DNA in Vector NTI.
- Examine and modify plot properties
- Modify plot layout
- Format sequence text
- Add new analyses to the display window

Follow the steps in the order shown. Figures show what your screen should look like at various points along the way.

1. In Vector NTI, Open a Molecule Display Window for pSV2-dhfr; Create a New Protein by Translating a CDS

For the first part of this tutorial, you will create a new protein in Vector NTI by translating a coding DNA sequence (CDS) found by using the ORF tool in a Molecule Display window. When you create the new protein, a new Molecule Display window opens from which you will then launch BioAnnotator. BioAnnotator can also be launched from Database Explorer or from the Start menu by selecting Start > Programs > Invitrogen > Vector NTI Suite 10 > Bioannotator.

Using techniques you learned in previous tutorials, in Vector NTI open a display window for pSV2-dhfr. To launch the ORF Setup dialog box, select Analyses > ORF or choose View > Display Setup, check the ORF box and click the ORF Setup button. Review the ORF settings in the ORF Setup dialog box and leave them as they appear. Click OK.
The ORF tool finds all open reading frames on the molecule according to the parameters set in the ORF Setup dialog box. ORFs are displayed in the graphics pane represented by arrows that indicate whether the ORF is found on the direct or complementary strand of the DNA/RNA molecule’s sequence. ORFs on the direct strand point clockwise; ORFs on the complementary strand point counter-clockwise.

Click on the arrow for the ORF that corresponds to the functional signal for the gene entitled “DHFR” (dihydrofolate reductase). The amino acid sequence translation from the selected ORF will form the basis for the new molecule.

Translate the ORF DNA to protein by selecting File > Create New Sequence > Using Translation Procedure (Protein) > Direct Strand. A New Protein Molecule dialog box opens with a default name of Translation of pSV2-dhfr for the new molecule. Change the name to DHFR and click OK. A new Molecule Display window for the new protein opens.

2. Launch BioAnnotator from a Vector NTI Molecule Display Window

In the new Molecule Display window, select Analyses > BioAnnotator > Analyze Selected Molecule. BioAnnotator opens featuring the new protein (Figure 15.2). If you had selected BioAnnotator > Open New Analyzer Window, BioAnnotator would have opened with empty panes, waiting for you to create a new project. Maximize the window by clicking the Maximize button ( ).
3. Review the BioAnnotator Display Window

The BioAnnotator interface is similar to those of other Vector NTI Advance applications: it consists of a menu bar, three toolbars that are context sensitive, a Text Pane, a Plot Pane, and Sequence Pane. Maximize the Display window and arrange its panes conveniently, using the split bars and scroll bars as you have in other tutorials.

The Pane selection buttons ( ), ( ), and ( ) are for quick navigation between panes.

Note: Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

In the Text Pane, open each of the folders with a double-click and review them. They contain descriptions of the molecule being analyzed in BioAnnotator. You'll note that dihydrofolate reductase is a small protein, 188 amino acids. The General Description folder verifies that the molecule was translated from a fragment of pSV2-dhfr.

Click in the Plots Pane to activate it. The Plot Pane displays visual results of physiochemical analyses as plots that you can manipulate as you have done with graphics in other display windows.

Activate the Sequence Pane by clicking on the Sequence Pane button ( ). If the sequence is long enough, you can navigate through the Sequence Pane by moving the vertical scroll bar. As in other Vector NTI applications, the cursor position in the Sequence Pane is noted on a pop-up label when the cursor is paused or on the Status Bar when you click in a certain position.

4. Examine and Modify Plot Properties

Click in the Plots Pane to activate it, then click on the Zoom In button ( ) two times, enlarging the plot. Move the horizontal split bar down to increase the Plots Pane viewing area. A plot's vertical axis shows the minimal and maximal values of analysis results; the horizontal axis displays either numerical positions in the sequence or residues, if the scale of the plot is enlarged sufficiently.

To review properties of any plot, click on it, then click the Properties button ( ), opening the Properties dialog box. Open this dialog box for the hydrophobicity plot. The Analysis tab displays the specific reference for this analysis, and the hydrophobicity index at pH 3.4 for each residue in the protein. Select the Datagram tab (Figure 15.3) and change the color to bright red. Click OK.

![Figure 15.3 Datagram tab of the Properties dialog box](image-url)
5. Modify Plot Layouts

Click the **Edit Layout** button ( ) on the toolbar. Click on the **Enable Overlapping** button ( ). Now plots can now be repositioned anywhere in the pane, even being superimposed upon each other. Left-click on the hydrophobicity plot and drag it on top of the hydrophilicity graph (Figure 15.4). Note how, in this case, the two graphs almost perfectly complement each other.

![Figure 15.4 Hydrophobicity plot superimposed on top of the hydrophilicity graph](image)

Drag the hydrophobicity plot back to its original position. Now move the cursor to one of the small black selection boxes at the bottom of the plot, where it changes to a vertical two-headed arrow. Drag the box down, modifying the plot height.

Turn off the **Enable Overlapping** option by clicking on the button again. With only the Edit Layout button still on, a plot can be moved, but dropped only in specific positions. Click and drag the lower plot to the top of the pane. When the cursor shape changes from to , a new position is allowed. When the cursor shape changes to , the new position is invalid. The upper pane now repositions itself in the lower position.

Click the **Edit Layout** button again to turn off that mode.

Select a region of the sequence on the plot using the click + drag technique. *Note that the same region is selected in the Sequence Pane.* Increase the scale of the selected region in the Plots Pane by clicking the **Zoom to Selection** button ( ). *Note that now you can read the individual residues in the Plot Pane.* To discern the plot data for a specific residue, pause the cursor over the residue position and read the popup label. Return to the original scale by clicking the **Fit to Window** button ( ).

6. Format the DHFR Sequence

Reposition the horizontal split bar so you can easily work in the Sequence Pane and activate the pane. Sequence text can be formatted in much the same way as you would in a word processor.
Select aa #131-140. Click on the **Format** button ( ), enabling the text format buttons you may be familiar with. Click on the **Bold** and **Italics** buttons (Figure 15.5).

![Figure 15.5 Formatting the DHFR sequence](image)

To reverse these actions, with the text still highlighted, click on the **Bold** and **Italics** buttons again. You can even change text color if you wish.

### 7. Open a DNA Project in BioAnnotator

You have launched BioAnnotator from a Vector NTI window with a protein molecule you created by translating DNA. Now let's open a BioAnnotator project containing a DNA molecule from the Vector NTI DNA/RNA database.

In the BioAnnotator window, select **File > Open**. Browse for the Demo Projects folder in the Open dialog box. Open the DemoProjects folder and click on DNA (or DNA.apf, if your computer is set to show file extensions). Click **Open** and BioAnnotator reads the file format and opens the display window for the DNA molecule, ColE1. Use the split bars to enlarge and optimize the display of the Plot Pane.

### 8. Add New Analyses to BioAnnotator

Since a DNA molecule is now open in BioAnnotator, the displayed default analyses have changed. Since an important part of configuring BioAnnotator is the selection of analyses to perform for a molecule, let's add some analysis to those already displayed. Click the **Analysis List** button ( ) to open the Analysis List setup dialog box (Figure 15.6).

![Figure 15.6 Analysis List setup dialog box](image)

The dialog box lists all of the analyses available for the molecule type open in BioAnnotator (DNA). Currently displayed analyses are listed in the Selected Analyses box. In the Available
Analyses list box, select **Melting Temperature (GC content)**. Click the **Add Analysis** button ( ) to move it to the Selected Analyses list. Using an alternative method, move **Nucleic Acid Distribution** to the lower list by double clicking on it.

The plot display order is determined by the order the analyses are listed in the Analysis List setup box. To modify the order, select Nucleic Acid Distribution and move up by clicking the **Move Up** button ( ). Click **OK** to generate the new analyses.

**9. Copy, Paste, and Print BioAnnotator Panes**

As in other applications of Vector NTI Advance, you can copy display window contents to the Clipboard or to a file. Make sure the Plot Pane is active. Position the plots for the display you would like to copy. Click the **Camera** button ( ) and select **All** and **Clipboard** in the dialog box that opens. Open a new word processing document and paste the Clipboard contents to it. The Camera feature copies and pastes pane contents exactly as you have specified and selected, even the colors in the copied pane.

The contents of any BioAnnotator pane can be printed the same as you do in other Vector NTI Advance Applications, by clicking the **Print** button ( ). To see a preview first, click the **Print Preview** button ( ).

**10. Close the Display Window and Exit Vector NTI**

Close BioAnnotator by selecting **File > Close, File > Exit**. You have finished your session with Vector NTI. Close the Molecule Display window using the Close command in the File menu. Quit Vector NTI by selecting **File > Exit**.
Introduction

ContigExpress is a program for assembling DNA fragments into longer contiguous sequences commonly called “contigs.”

Work done in ContigExpress is saved as a ContigExpress Project, a file with a .cep extension in which you can store fragments, their assemblies, assembly options, and descriptions related to your project. In ContigExpress, fragments can be edited directly, with the chromatograms in full view. Changes are tracked and a history is maintained. The contigs thus generated can then be saved, exported into other applications of Vector NTI Advance or exported to third-party tools available on the Internet.

This tutorial can be completed in one session or divided into three sessions as noted. At the end of this tutorial, you will be able to:

- Create a ContigExpress project
- Navigate and manipulate the CE Project Explorer window
- Assemble contigs
- Trim or edit fragments in the Fragment Viewer
- View and edit fragments and contigs in the Contig Viewer

Follow the steps of the tutorial in the order shown. Figures show what your screen should look like at various points along the way.

Session 1 Project Explorer

1. Open a ContigExpress Window

Launch ContigExpress (CE) from the Start menu by selecting Start > Programs > Invitrogen > Vector NTI Advance 10 > ContigExpress. You can also launch ContigExpress from the Assemble menu in most other Vector NTI Advance applications. ContigExpress opens with an empty Project Explorer.
To open the project, in the Project Explorer window, select **Project > Open Project**. Browse to `C:\Program Files\Invitrogen\Vector NTI Advance 10` (if a default installation has been performed). Open the Demo Projects folder and double-click on DemoProject.cep to open the project in the Project Explorer window. The ContigExpress DemoProject contains 13 fragments, listed in the List Pane of the Project Explorer (Figure 16.1):

![Figure 16.1 Fragments listed in the List Pane of the Project Explorer](image)

Many files saved from an automatic sequencer have an `.abi`, `.ab1` or `.esd` extension. Add one `.abi` format fragment to the list by selecting **Project > Add Fragments > From ABI file…**

This option opens the Import Sequence From dialog box (Figure 16.2). Select Rb.abi and Sample 1303.abi and click **Open**.

![Figure 16.2 Import Sequence from dialog for browsing in one or more sequence files](image)

Often the names of the imported fragments registered within the fragment file do not directly correspond to their Windows file name designations. ContigExpress can either assign Windows file names or fragment names that are found inside the imported files to the newly imported fragments. Upon pressing the **Open** button, ContigExpress asks you to specify your naming preference (Figure 16.3):

![Figure 16.3 Naming preference dialog](image)
Figure 16.3 Indicate the naming preference of imported sequence files

If you choose No, ContigExpress imports the fragment, naming it according to the Windows file name, Rb and Sample1303. For this example, however, choose Yes. The files Rb.abi and Sample1303.abi are imported and named according to their internal file names and are therefore listed as “RB 1” and “xb-control” in the ContigExpress List Pane.

After you specify your naming convention preference, a dialog box appears, confirming the number of fragments imported and time elapsed during import.

3. Examine the ContigExpress Project Explorer Window

As noted, ContigExpress opens to an empty Project Explorer window to which you added a list of molecules. The window is similar to other Vector NTI Advance windows with a title bar, menu bar and tool bar (Figure 16.4). The Tree Pane on the left and the List Pane on the right are divided by a movable split bar. A Status Bar at the screen bottom displays the Project Explorer status.

Note: Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

4. Create Contigs

Start Assembling

To start an assembly, select all the fragments except an arbitrary one, say, xb-control, by SHIFT + CLICK on all the names on the list except that one (Figure 16.4). Click the Assemble Selected button ( ). This button is enabled only when two or more fragments are selected.
**Note:** If you want to select all the sequences in the Project window for assembly, right click on the first sequence in the list and choose **Select All** from the shortcut menu.

An Assembly Progress dialog box allows you to monitor the assembly process or cancel the assembly at any time. After the assembly process is finished, a summary dialog box displays, showing the elapsed assembly time, the number of contigs assembled, and number of fragments used.

**Assembly Setup**

More than likely you would want to control the assembly through fine-tuning the assembly parameters, rather than relying on the default settings. To see what assembly parameters are available for control, choose **Assemble > Assembly Setup** from the menu or choose the **Assembly Setup** button on the toolbar ( ). The Assembly Setup dialog box (Figure 16.5) appears (more detailed information about Assembly Setup can be found in Chapter 30). For now, proceed with the assembly you created using default values. Click Cancel on the Assembly Setup dialog and return to the Project Explorer.

![Assembly Setup dialog box](image)

**Figure 16.5 The Assembly Setup dialogs**

**Examine Assembly Results**

Once you have performed the contig assembly, the Tree Pane in ContigExpress Project Explorer shows the new assembly, Assembly 1 (Figure 16.6).

An assembly can be made up of one or more contigs. Click on Assembly 1 to select it. All of the project fragments and contigs are listed in the List Pane. Assembly 1 includes two assembled contigs, with Contig 1 made from eleven fragments and Contig 2 composed of two. Unassembled fragments are shown below the contigs. Assembly 1 includes one unassembled fragment, RB 1, that was selected for assembly, but did not assemble (indicated by the grey arrow icon to
the left of its name) and one fragment, xb-control, which was not selected for assembly (Figure 16.6). Note that many of the fragments that are included in contigs underwent trimming at their 5' or 3' ends or both. This results from the operation of the assembly algorithm, although it can be turned off if so desired (Figure 16.6).

Let’s create another assembly, a “child” of Assembly 1. For this, hold down the Shift key and click on fragments ONE17KANR and ONE9KANR, thus selecting seven fragments in Contig 1: ONE17KANR, ONE13KANR, ONE15KANF, ONE10KANF, ONE11KANR, ONE16KANF, and ONE9KANR (Figure 16.7). Click the Assemble Selected button as before.

Another assembly derived from Assembly 1 is listed in the Tree Pane (Figure 16.8). The numbering system (Assembly 1.1) describes its relationship to the parent contig. Note that Contig 2 has been inherited unchanged from the parent assembly, Assembly 1, into Assembly 1.1. Also note that in Assembly 1.1, the four unselected fragments, ONE2KANR, ONE3KANR,
ONE6KANR and TWO11KANR, are “released” to the pool of unselected fragments as indicated by the icon.

Figure 16.8 “Child” Assembly 1.1, derived from fragments in Contig 1 from Assembly 1

5. Review and Modify Viewing Options

By default, the Project Explorer is configured to display five columns listing the identification code of each fragment, their original lengths, current lengths, and 5’ and 3’ trims. Let’s modify the viewing options to display more information for each fragment. Select View > Options to open the Options dialog box (Figure 16.9):

This dialog box, similar to those in other Vector NTI Advance applications, has three tabs. The Details View tab controls the information displayed in the List Pane of Project Explorer. The Status Bar tab controls the information displayed on the Status Bar for a selected item in the
Project Explorer. The **Copy Fragment/Contig** tab controls the copy format for fragments and contigs.

On the **Details View** tab, the top to bottom order of the Displayed Fields corresponds to the left to right order of columns in the List Pane. In the **Available Fields** pane, select **Non-ambiguous Bases** and click the >> button to add it to the list of displayed fields. In the **Displayed Fields** pane, select **Non-ambiguous Bases** and click the **Move Up** button to move it to the top of the list.

Click **OK**. Note that **Non-ambiguous Bases** and **Length** now appear as columns to the immediate right of the **Name** column in the Project Explorer List Pane (Figure 16.10):

![Figure 16.10 Non-ambiguous bases field is added to the List Pane display](image)

On the toolbar, the List Format group of buttons control the list display (Figure 16.11). Click each of these buttons and observe how the information display of the Project Explorer changes.

Leave the **Details** button ( ) selected.

![Figure 16.11 List Format group of buttons](image)

The next group of buttons allows you to filter data displayed in the List Pane (Figure 16.12):

![Figure 16.12 Filter Data group of buttons](image)
By default all 4 toggle buttons, or viewing filters, are selected, showing the maximum amount of information. To see how they work, make sure that Assembly 1.1 in the Tree Pane is selected. Note that no matter how you display the list, all elements are still identified by one of the following icons:

- Contigs ()
- Fragments included in the assembly ()
- Fragment not included in the assembly ()
- Fragments not considered for the assembly ()

Now click the **Show Contigs** button (), turning it “off”. Contigs 2 and 3 and their fragments are no longer displayed in the List Pane (Figure 16.13):

![Figure 16.13 Hide contigs control leaves only non-assembled fragments in display](image)

Click this button again to restore the original display, then click **Expand Contigs** button ( ), turning it “off”. Now the fragments included in Contigs 2 and 3 are not displayed (the contigs are not “expanded”).

Click the **Expand Contigs** buttons again to restore the original display. Click each of the remaining two buttons, turning them off then on again in turn: the **Show Unassembled Fragments** button () and the **Show Other Fragments** button (). Note which fragments are excluded and which included in each display. Finish with all filters enabled (all buttons pushed) as it was in the original display.

You can use any combination of enabled and disabled data filters for list display.
6. View Contents in the Tree and Content Panes

The default enabled viewing mode in the Tree Pane is called the History View. It shows the historical relationship between the assemblies in a tree form. For a view with a different perspective, click the Content View button ( ).

The Content View (Figure 16.14) does not display the mutual relationships of the assemblies, but it does enable you to open each assembly from the Tree Pane, displaying each assembly’s contents in the List Pane. Click the History View button ( ) to return to History View, which displays assembly relationships (trees), but not their contents, in the Tree Pane.

7. Manage Fragment Lists and Project Items

ContigExpress can use the Windows clipboard for standard copy and paste operations. Most operations are accessible through the Edit and shortcut menus, or dedicated toolbar buttons.

Select fragment xb-control and click the Copy button ( ) (enabled when anything is selected in the List Pane). Click the Paste button ( ) and Copy of xb-control appears in the List Pane.

Project Explorer allows you to change the names of the project items and delete them from the project. Select the fragment Copy of xb-control and click the Rename button ( ), enabling the text box (Figure 16.15). In the text box, change the name to 123 and press ENTER, renaming the fragment.

Select 123 and click the Delete from Project button ( ). Confirm the action. Fragment 123 is deleted from the project and the Explorer is reset to the upper level of the project (Fragments (MAIN)).
Select **Assembly 1.1** and fragment **ONE4KANR** in Contig 2 of this assembly. Click the Delete from Project (❌) button. ContigExpress warns you that the selected fragment is included in several assemblies and deleting this fragment will delete those assemblies as well (Figure 16.16).

![ContigExpress warning](image)

**Figure 16.16** Deletion of a fragment will dismiss assemblies in which the fragment is in a contig

Click **No**, and the deletion will not continue. Had you clicked **Yes**, all assemblies would have been deleted.

8. **Save and Close a ContigExpress Project**

Even though you may wish to continue the tutorial, save and close the project and then open it again to become familiar with these functions.

To save the project, select **Project > Save As**. Name the project **Tutorial Project** and select the Demo Projects folder as the destination.

To close your project, select **Project > Close Project**. ContigExpress reverts to its initial state, with an empty workspace. You can either start a new project by importing the fragments or open an existing project from the disk.

Exit ContigExpress as you would close any other Windows application.

**Session 2 Working in Fragment Viewer**

1. **Open and Inspect the Fragment Viewer**

Open the Project Explorer and select Project on the menu bar. The name of the project you worked with last is shown in the used projects list at the bottom of the drop down menu. Select Tutorial Project, opening it in Project Explorer. Select the upper level of the project (Fragments (MAIN)) and in the List Pane, double-click the fragment **ONE8KANR** to open it in a Fragment
Viewer (Figure 16.17). Maximize the window and arrange the panes conveniently by dragging the split bars with the mouse.

![Fragment Viewer Screenshot](image)

**Figure 16.17 ContigExpress Fragment Viewer**

The Fragment Display window consists of a title bar, a menu bar, three toolbars (Main Toolbar, Window Toolbar, and Pane Toolbar) and three panes: a Text Pane, a Sequence Pane and a Chromatogram Pane. The panes are divided by movable split bars.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

**Text Pane**

Click the **Text Pane** button ( ) to activate it. The Text Pane lists properties of the fragment, such as its name, general description, some record of the chromatogram data, and feature annotations on the fragment, if any.

Click the + to the left of the General Description line to open and view the folder contents. Double click on the folder icon to close it.
Sequence Pane

Click the **Sequence Pane** button ( ) to activate it. By default, only one strand of the sequence is shown. Click the **Show Two Strands** button ( ) from the toolbar or **View** menu. Now both strands of the sequence are displayed (Figure 16.18).

![Figure 16.18 Double-stranded sequence display](image)

Scroll to the end of the sequence. Note that 2 bases (CA) were dropped down from their original row, and replaced by down-pointing arrows. This is a result of clipping (default - on) performed by the assembly process. Activate the Chromatogram Pane and scroll to the far right to see how the clipped bases are represented in the chromatogram. Return to the Sequence Pane by clicking it.

To modify the number of sequence blocks displayed per line, click the **Lane Width** button ( ) and enter 15 in the Blocks per Line field of the Set New Number of Blocks dialog box. Click **OK** and the number of blocks per line increases as directed (Figure 16.19).

![Figure 16.19 Sequence Pane displayed as 15 blocks (5 bases each) per line](image)

The same control is available via the **View > Grouping > Set Number of Blocks** menu. Change the blocks display back to 10.
Chromatogram Pane

Click the Chromatogram Pane button to activate the pane (Figure 16.20). The vertical scroll lever on the left side of the Chromatogram Pane can be used to resize chromatogram peak height.

Individual base traces can be displayed or removed by clicking the corresponding Show Trace toggle buttons on the toolbar or by choosing View > Show Traces for the particular base you want to show or hide. For example, to hide the chromatogram trace for Ts only, make sure that the Show Trace -T button is not pressed, but that the Show Trace A, C and G buttons are pressed. Press each button to see the effect on the Chromatogram Pane.

2. Sequence Selection Techniques

In ContigExpress, positioning the cursor in the Sequence Pane transforms it into a short, vertical, blinking I-beam called a caret ( ). When a sequence is selected and highlighted, one end (the “active” end) is marked with a caret. The current position of the caret is always shown in the Position Box on the status bar (Figure 16.21):
Click the position box to open a Set Caret Position dialog box that can be used to position the caret anywhere in the sequence. Enter the nucleotide position number and click OK.

For setting a selection range, click on the Selection box on the Status Bar, opening the Set Selection dialog box. Set the selection from 10 bp to 30 bp and press OK.

The selection is simultaneously highlighted in the Sequence and Chromatogram Panes and shown in the Selection Box on the status bar (Figure 16.22).

To modify the 5' and 3' ends with the mouse:

- Click the 5' end of the selection, moving the caret [note its change to a 5' mark (↑)] to that position and activating the 5' end. CLICK + DRAG the 5' end to a position near the start of the molecule. Observe the selection box as you drag. The caret remains with the end of selection you are dragging.

- The end of the selection marked with the caret can also be moved with arrow keys on the keyboard. Hold the SHIFT + the left or right arrow keys to move the 5' end of selection one nucleotide at a time. (If you hold down CTRL + SHIFT + use the arrow keys, the selection is changed 5 nucleotides at a time.) When the 5' end is placed on the start of the molecule (1 bp), release the SHIFT key.

- To set a selection with the mouse, click anywhere inside the sequence and drag the caret with the mouse, releasing when the selection is complete.

Note that the selected fragment is also displayed concurrently in the Chromatogram Pane. This occurs no matter where the selection was made.

3. Display ORFs and Translate the Nucleotide Sequence

Display ORFs for the Sequence

In the Fragment Viewer Sequence Pane choose Show ORFs from the View menu or click the Show ORFs button ( ) on the toolbar. ORFs are displayed for the direct and complementary strands (Figure 16.23). If single-stranded sequence is displayed, only the ORFs for that strand.
are shown. ORF finding criteria (e.g. minimum length, inclusion/exclusion of stop codon, etc.) can be adjusted via the **View Options** dialog **ORF** tab from the **View** menu or toolbar button.

![Figure 16.23 ORF display in Sequence Pane](image)

**Figure 16.23 ORF display in Sequence Pane**

**Translate the Nucleotide Sequence**

Select the region from 127 to 138 bp. The Chromatogram Pane has also made the corresponding selection and has scrolled its graph to the selection. Click the **Translate Selection** button on the **View** menu or toolbar. ContigExpress shows the amino acid abbreviations above the selected sequence (Figure 16.24).

![Figure 16.24 Direct strand translation of a selection in Sequence Pane](image)

**Figure 16.24 Direct strand translation of a selection in Sequence Pane**

145
To translate the complementary strand of the selection, click the **Translate Complementary** button ( ) on the menu or toolbar. If the sequence is displayed as single-stranded, the pane automatically switches to two-strand mode to show the complementary translation below the selection (Figure 16.25):

![Figure 16.25 Two-strand translation of a selection in Sequence Pane](image)

To toggle between 1 and 3 letter codes in the translations click the **Show 1-letter AA codes** button ( ) on the **View** menu or toolbar.

To delete all translations, click the **Clear All Translation** button ( ) on the **View** menu or toolbar. Click the **Show Two Strands** button ( ) to return to 1-strand display. Notice that when you return to single-strand sequence display, the ORF display changes to direct ORFs only.

Click on the **Show ORFs** button ( ) to remove the ORFs display.

### 4. Edit the Nucleotide Sequence in the Sequence Pane

You may freely edit a fragment in the Fragment Viewer only if it is not part of a contig. A fragment that is part of a contig, as ONE8KANR currently is, opens in Read-only mode. To edit such a fragment, you must open the contig in a Contig Viewer and edit the fragment there. (Any edits made there will be saved to the fragment.) Otherwise, you must dissolve the contig before you open and edit a fragment contained in it.

In Project Explorer, select Contig 1 from Assembly 1 and select **Dissolve Contig** from the shortcut menu. This discards Contig 1 and frees up member fragments for editing; repeat on Contig 3 of Assembly 1.1. Double-click the fragment, ONE17KANR to open it in Fragment Viewer. Fragment ONE8KANR remains in Read-Only mode and Contig 2 is unaffected, since they are not descended from the dissolved contig.
You will learn three ways to edit the fragment sequence. All are illustrated in the same screen shot, Figure 16.26.

1. **Nucleotide Deletion** In the Sequence Pane, select nucleotides 316 through 319. Press the Delete key on your keyboard. The selected symbols move under the row, with downward pointing arrows (↓) taking their place.

2. Now move your attention to the Chromatogram Pane. There is a sequence above and one below the graph. Before we started editing, they were identical. The sequence above the graph reflects all your editing commands, the sequence below always remains in its original form, thus giving you a useful reference to track and compare your changes. Note that the deletion resulted in each deleted nucleotide being replaced by a dash (-) in this sequence.

3. **Nucleotide Insertion** Back in the Sequence Pane, move the caret in the Sequence Pane to position 307 and type TTT. Three Ts with upward pointing arrows (↑) below them are inserted at the caret position.

4. In this case, the arrows indicate that symbols were inserted in a sequence (“added to” the sequence) into a position, not substituted for other nucleotides. The sequence above the graph in the Chromatogram Pane also shows the insertion (in another color). Note that a gap appeared in the graph because there is no data to build chromatograms for newly inserted symbols.

5. **Nucleotide Substitution** Select nucleotide 303 and type G. The selected base is replaced by G, and A is shown below G to help you track the change. The chromatogram does not change because the G was not inserted; it just replaces the original base call. You can only replace the bases one by one.

By default, peaks for deleted bases are still shown in the chromatogram. One consequence is that although bases have been deleted, they are still considered part of the sequence during translation. Deleted peaks can be removed using the **Show Deleted Peaks** feature, allowing translation in the new/resulting reading frame. Press the **Show Deleted Peaks** button or choose **View > Show Deleted Peaks** to turn off the Show Deleted Peaks function. When a message appears indicating the fragment will switch to read only mode, choose **Yes**.

Make a selection across the area where you have edited bases from 301 to 375 bp and trans-
late the sequence using techniques previously described. Notice the translation considers the deletions in the sequence and proceeds according to the new frame (Figure 16.27).

Click the **Show Deleted Peaks** button again to reactivate the function. Notice the translation now includes the 4 deleted peaks, as if no deletions have been made (Figure 16.28). Consequently, the downstream translation is changed.

Erase your translation using techniques described earlier. You can use the **Undo** and **Redo** commands to reverse or restore your editing actions. Click the **Undo** Main toolbar button several times, and watch your earlier edits disappear. Eventually it becomes disabled and both Sequence and Chromatogram panes return to their original sequences.
5. Edit a Sequence in the Chromatogram Pane

Activate the Chromatogram Pane by clicking the Chromatogram Pane ( ) Toolbar button. In the top, editable sequence in the Chromatogram Pane, select “T” at bp 11 and type C. As in the Sequence Pane, T is replaced by C. Note that each of these actions is simultaneously reflected in the Sequence Pane.

Play with editing in both panes to become familiar with all features, but do not be afraid to edit too much. You can restore the original sequence by selecting Fragment > Revert to Saved. Note that all editing operations are reflected in both the Sequence and Chromatogram Panes. They can be summarized as follows (NTs = nucleotides):

<table>
<thead>
<tr>
<th>Action</th>
<th>How to Perform</th>
<th>Sequence Pane Result</th>
<th>Chromatogram Pane Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete</td>
<td>Select residues; press Delete</td>
<td>Replaces NT with (↓); original NT moves below strand</td>
<td>(- - - -) Appear in upper, editable sequence</td>
</tr>
<tr>
<td>Insert</td>
<td>Place caret; type new NTs</td>
<td>(↑) appears below new NTs; new NTs are colored</td>
<td>A break appears in the chromatogram</td>
</tr>
<tr>
<td>Replace</td>
<td>Select existing NTs and type new NTs one by one</td>
<td>New NT appear in sequence; replaced NTs moved below strand</td>
<td>New NT appears in upper sequence; no break or change in chromatogram</td>
</tr>
</tbody>
</table>

Table 16.1  Editing operations for Sequence and Chromatogram Panes

6. Close the Fragment Viewer

Select Fragment > Close or click the close box to close the Fragment Viewer. If asked, confirm the operation.

7. Exit ContigExpress

Close ContigExpress by selecting Project > Close Project. When asked whether to save the changes, click No so you can continue to use the project with the assemblies and contigs created (and dismissed during this session) in the first part of this tutorial.

Session 3 Working in the Contig Viewer

When you are in a ContigExpress project, selecting and opening a fragment brings up Fragment Viewer. Selecting and opening a contig opens a Contig Viewer.

1. Open and Examine a Contig Viewer

Re-open ContigExpress as you have done before. Load the Tutorial Project that you created in Session 1 by selecting Project > Tutorial Project.cep. In the Project Explorer window, select Assembly 1 and double-click on Contig 1 in the List Pane, opening the Contig Viewer (Figure 16.29). Maximize the window.
Contig Viewer has many elements similar to the Fragment Viewer: a title bar, three toolbars and three panes: a Text Pane, a Graphics Pane and an Alignment Pane. Arrange the panes conveniently by dragging the split bars with a mouse. A Status Bar at the screen bottom displays the Project Explorer status.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

![ContigExpress Contig Viewer](image1)

**Text Pane**

To activate the Text Pane, click on the **Text Pane** button ( ). The Text Pane is similar to other Text Panes with its folders containing various properties of the fragment.

Open the Fragments Folder and select fragment **ONE11KANR**. When you click on it, this fragment is located in the remaining two panes: in the Graphics Pane it is highlighted in blue, and in the Alignment Pane, its name is highlighted (Figure 16.30).

![Fragment selection in Contig Viewer is synchronized in all 3 panes](image2)
It is possible to perform additional assemblies from the Text Pane without returning to the Project Explorer. **Assemble Selected Fragments** becomes accessible from the **Assemble** menu or a toolbar button ( ) when two or more sequences are selected. The **Assemble Not Selected** button (toolbar only) becomes available when there are at least two unselected sequences and one selected sequence in the Text Pane. The **Assemble Not Selected** command is similar to **Assemble Selected Fragments**, except that it excludes the selected fragment(s) when attempting to assemble a Contig.

When either of these commands is executed, the new assembly is performed and added to the project as a child assembly of the assembly that the current Contig is a part of. For example, if the Contig is part of Assembly 1, the new assembly will be called Assembly 1.1.

**Graphics Pane**

Click the **Contig Graph** button ( ) in the Window toolbar to activate the Graphics Pane.

By default, the data in this pane is drawn in **Fit to screen** mode. The Graphics Pane contains horizontal arrows representing the relative positions of the fragments forming the contig. The arrowheads indicate the direct or complementary placement in the contig of the respective fragment.

Below the fragment arrows are a Coverage Map and a scale, or Weight Graph, with position markers for this contig. The Coverage Map summarizes the depth of coverage and directionality of the fragments. The Weight Graph displays the quality of the assembly. Click the Zoom In button until a horizontal scroll bar appears below the weight graph and position markers are shown every 50 bases. Move the scroll bar on the Weight Graph so that you can view the region at about position 1235. Note the green vertical line there. Position the cursor at that site and click the **Zoom In** button ( ) several times to enlarge the graph. You may need to scroll the horizontal scroll bar to keep the site you marked in view as you enlarge the area. **When the maximal zoom is reached, the Zoom In button becomes disabled.**

![Figure 16.31 Using the Weight Graph to identify positions of poor consensus](image)

The negative peak in the Weight Graph line (Figure 16.31) indicates a position where the consensus is poor. In the Alignment Pane, look at the symbols at the same position. The nucleotide at bp 1235 appears to be unresolved. Instead of having to thoroughly examine all nucleotides...
sequences throughout the contig, you can see the “bad” points with just one glance at the Weight Graph.

For details regarding the weight calculation, refer to Chapter 31.

Click **Fit to Screen** from the View or shortcut menu, or toolbar button ( ) to return to the standard sizing.

You can make selections in the Contig Viewer using techniques you have already learned. Select nucleotides from 2000 to 2077 bp by dragging the mouse on the Graphics Pane. Note that the selection is concurrently scrolled to in the Alignment pane, with blue vertical lines at the start and end coordinates of this selection (Figure 16.32).

![Figure 16.32 Sequence selection in the Contig Viewer](image)

ORFs can be displayed in the Graphics Pane of the Contig Viewer by activating the Graphics Pane and either choosing **View > Show ORFs** from the menu or clicking the **Show ORFs** button ( ) on the toolbar. Click the **Show ORFs** button. The ORFs display area appears below the fragment display in the Graphics Pane (Figure 16.33). Green bars in the ORFs display area represent individual ORFs. Click on a green bar to highlight the ORF in the Graphics Pane. The
selected ORF is also scrolled to in the Sequence Pane; note the blue vertical line at the end coordinate of this selection.

Gaps that appear in the contig consensus are ignored during ORF calculation, i.e., they are skipped over. Additional parameters for the ORFs search can be configured in the Contig View Options dialog box. Choose View > View Options from the menu in the Contig Viewer or click the Viewing Options button ( ) on the toolbar to launch the dialog box. Select the ORF tab on the Contig View Options dialog box (Figure 16.34).

Acceptable start and stop codons can be designated here. Minimum ORF size for the search can also be set, as well as the option to show nested ORFs. Close the Contig View Options dia-
log box by clicking on the Cancel button. Click the Show ORFs button to remove the ORFs from the Graphics Pane display.

Activate the Graphics Pane and choose View > Show Contig Coverage from the menu or click the Show Contig Coverage button ( ). The graphical representation of how the assembled fragments cover Contig 1 is removed from the Graphics Pane. Click the Show Contig Coverage button again to return the graphic to the display.

2. Review the Contig in Contig Alignment Pane

Click the Alignment button ( ) in the Toolbar to activate the Contig Alignment Pane (Figure 16.35). The Alignment Pane displays the nucleotide sequences of the contig consensus as well as those of the fragments forming the contig and their positions. In the standard configuration of this Viewer, positions/columns in the consensus where the fragments do not completely agree are highlighted in yellow. The particular bases at such positions that disagree with the consensus appear in red.

![Figure 16.35 Contig Viewer – Alignment Pane with ambiguous positions highlighted](image-url)
Let's study the elements of the Alignment Pane (Figure 16.36):

- The **fragments name list** on the left contains all fragments used in this contig. Reorder fragments in this list by selecting and dragging vertically.
- The **scale** gives a reference for all fragments and the consensus shown in the alignment.
- The **fragment current position** is the relative (counted from the fragment start) position of the currently shown left end of the respective fragment.
- The **fragment sequences**, with the option of showing chromatograms, if available. Some fragment ends were clipped as part of the assembly process; hovering the mouse alongside such a fragment end will display the trimmed sequence in a tooltip.
- The **consensus sequence (nucleotide format)** shows the nucleotide consensus for the assembly.
- The **consensus sequence (symbol format)** shows the symbol consensus for the assembly. In the symbol consensus, blank spaces indicate identical residues, dots (*) indicate positions with an IUPAC ambiguity symbol in a fragment, and plus symbols (+) indicate ambiguous positions containing specific base disagreements and/or gaps. The two consensus sequence tracks, nucleotide and symbol, may be repositioned at the top of the Alignment Pane, using the Alignment on Top command from the shortcut menu.

You may navigate easily among ambiguous positions using Find Next/Previous from the View menu or toolbar button ( ).

There are two Find features in the Contig Viewer: Find Fragment and Find Sequence. Find Fragment can be be activated from the Edit menu in any pane. In the dialog (Figure 16.37), enter the name of a fragment and click Find. Wild cards (*) are permitted in the search name...
(e.g. *16*). If a result is returned, double-click on it and the Viewer will jump to that fragment in the Text and Graphics panes.

![Find Fragment dialog](image1)

**Figure 16.37 Find Fragment dialog**

**Find** can be used to find a specific sequence in the Alignment Pane. Click on the **ONE6KANR** fragment name in the Alignment Pane list to select it. Place the cursor at the beginning of the sequence. Launch the Find sequence dialog box either by choosing **Edit > Find** from the menu or by clicking the **Find** button ( ) on the toolbar. When Find dialog box appears, type in the sequence “gagactccatt” (Figure 16.38). Click the **Complementary Search** checkbox.

![Find (sequence) dialog](image2)

**Figure 16.38 Find (sequence) dialog**

Click the **Find Next** button. The Alignment Pane jumps to and highlights contig positions 2470 to 2480 where the complementary sequence has been found (Figure 16.39). The **Set Selection** box on the Status bar also reports the position of the found sequence using the fragment coordi-
nates (128-138). The search is case-insensitive and can operate on the consensus as well as individual fragments. Click **Cancel** to close the Find dialog box.

Figure 16.39 Find a sequence in the Alignment Pane

Click on the **ONE6KANR** fragment name in the Alignment Pane. Right-click anywhere in the Alignment Pane, opening the shortcut menu. Select **Show Chromatogram for ONE6KANR**. The chromatogram for the selected fragment is shown below its sequence, the same as it was in the Chromatogram Pane of the Fragment Viewer (Figure 16.40).

Figure 16.40 Chromatogram display in Alignment Pane
Right-click anywhere in the Alignment Pane. Select **Show All Chromatograms** from the shortcut menu. Now the chromatograms for all four fragments at this position are displayed (Figure 16.41):

![All chromatograms displayed in Alignment Pane](image)

**Figure 16.41 All chromatograms displayed in Alignment Pane**

Right-click anywhere in the Alignment Pane and choose **Hide All Chromatograms** from the shortcut menu. Chromatograms in the Alignment Pane are no longer displayed.

The contig consensus sequence can be translated in the Alignment pane in any one or all of the three direct frames. Translate the contig sequence in the first frame by activating the Alignment Pane and choosing **View > Consensus Translation > 1st frame** from the menu or by pressing the **Consensus Translation 1st Frame** button ( ![button] ) on the toolbar. Translate the contig sequence in the second and third frames by choosing their respective menu commands or toolbar buttons (Figure 16.42). Gaps in the consensus are ignored in the translation, with the result being that the introduction of gaps does not result in frameshift. Toggle between viewing transla-
tions in single-letter or three-letter codes using the **Show 1-letter AA codes** button on the toolbar.

![Figure 16.42 Amino acid translations of the contig consensus](image)

Translutions can be removed by clicking the **Consensus Translation** buttons again or by deselecting the translation options from the **View > Consensus Translation** menu.

**3. Edit a Fragment in the Alignment Pane**

Fragments can be edited in the Alignment pane of the Contig Viewer. Highlight the fragment name **ONE9KANR** in the Alignment Pane and open the **Set Selection** box on the Status bar. Enter coordinates 100 to 110 and check **Fragment radio button** and click **OK**. This jumps the Alignment Pane to a sequence TGTAAAATCCC at positions 1923-1933 in the contig. The fragment coordinates, read 5' to 3', are 611-621, as you can read on the status bar. (Coordinates 100-110 refer to the coordinates as the original fragment is positioned in the contig – in this case, a complementary positioning.)

Click on the "Use Enhanced Edit Mode" button (/fixtures/toolbar.png) on the toolbar to allow editing. Choose **Edit > Delete Sequence from 611 to 621 bp** from the menu. Because this fragment is a part of a contig in Assembly 1.1, in addition to the current assembly, the following message appears (Figure 16.43):

![Figure 16.43 Approval for fragment editing that may dissolve other contigs](image)

Click **Yes** to accept the changes. When you proceed with the deletion, the bases are deleted from the chromatogram and the resulting gap is closed. A vertical red bar appears in the sequence marking the position where the deletion was made (Figure 16.44). The consensus
and the alignment are changed radically because sequence originally to the right in ONE9KANR is shifted left, and is out of alignment with the other fragments in this region.

Figure 16.44 Deleting a fragment sequence in the Contig Viewer Alignment Pane

If you save the edited contig and open the edited fragment in the Fragment Viewer, the editing changes are reflected in the fragment sequence.

Undo/Redo functions can be used to revert or repeat the changes you make. Click the **Undo** button until it becomes disabled, and the fragment and the consensus return to their original state.

Highlight fragment **ONE13KANR**, select bp 208 G and type T. If asked to confirm the edit, click the **Yes** button. The G is replaced in the fragment and the contig consensus is updated (Figure 16.45). In the fragment, the new base is colored teal to indicate it is a replacement.

Figure 16.45 Replacing a fragment sequence in the Contig Viewer Alignment Pane

Additional editing operations in the Alignment Pane include sliding entire fragments into place manually and sliding a block within a fragment into an adjacent gap. Use the **Move Fragment** buttons ( ) and **Move block to the left/right** buttons ( ), respectively. Find and select appropriate fragments/regions and try them.
The only limitation of fragment editing in the Contig Viewer is that if it affects contigs in other assemblies containing the newly edited fragment, those contigs must be dismissed.

Fragment in-place editing and update of the consensus on the fly makes the Contig Alignment pane a convenient tool for working with contigs after their initial assembly. Remember that after editing is performed in the Contig Viewer, fragments can be realigned as needed by reassembly after selecting the appropriate fragments in the Text Pane.

At any time, before or after editing, you can generate a Difference Report comparing any two fragments, or a fragment and the consensus. Select ONE13KANR in the fragment name list and check Compare To from the View or shortcut menus. In the dialog, accept the default comparison to the consensus and click Compare. The resulting report (Figure 16.46) may be saved as a .txt file to your computer.

When editing has been performed in a Fragment/Contig Viewer, separate prompts will appear upon closing both the Fragment/Contig Viewer AND the ContigExpress Project Explorer requiring approval for the edits to be saved.

4. Exit ContigExpress

You have finished the ContigExpress tutorial. Close ContigExpress as you would close any other Windows application.
Introduction

The Vector NTI database is a collection of molecules and laboratory data, organized for easy retrieval and management in Vector NTI’s Database Explorer. Eight different types of objects are stored and organized in databases and subsets in Database Explorer:

- DNA/RNA molecules are annotated with Feature maps. Upon import from other sources, molecule data is parsed and stored in an internal format. Users can add molecules to the database through import or creating basic or constructed molecules.
- Protein molecules are also annotated with Feature maps. Like DNA molecules, upon import from other sources, protein molecule data is parsed and stored in an internal format. Users can add molecules to the database through import or creating basic molecules.
- Restriction enzymes are imported from the REBASE database; data for the enzymes in the Vector NTI database is parsed and stored in an internal format. Users can add other restriction enzymes from the REBASE file included in the Vector NTI software.
- Oligonucleotides in the database are user-defined except for several included in the Vector NTI software for demonstration purposes.
- Gel markers commonly used in molecular biology laboratories are included initially in the Vector NTI software, but new gel markers can be created by the user.
- Citations for physiochemical analysis of molecules in BioAnnotator and other Vector NTI analyses are stored in the Vector NTI Database Explorer.
- BLAST results can be stored independent of molecules in a BLAST Results database.
- Analysis results performed on DNA/RNA and Protein Molecules, such as PCR analysis, PFAM analyses, etc. can be stored in the Vector NTI Database.

Very similar in functionality to the Windows 95/98/Windows NT interface, Database Explorer supports intuitive browsing of local and shared Vector NTI databases, clipboard data exchange, drag and drop operations and other functions typical of window-based database management.
From the Exploring Database window, you can perform the following:

- Create new molecules
- Sort, Edit and delete molecules
- Search the database for text sequence, motifs, feature types, keywords etc.
- Organize our data into convenient groups (subsets)
- Import and export data (this is inactive in the demo version)
- Open the other applications of Vector NTI Advance—AlignX, BioAnnotator and Con
tigExpress
- Copy and save or print a list of the objects in a subset.

Parent-Descendant relationships (to keep track of your constructs), user fields, comments, key-
words etc., are kept for all molecules in the database

All database molecules and other objects can be placed into “archives”—data files of special for-
mat—that can be transferred to another computer (Mac or PC) and read by Vector NTI there. 
Through archives, you can share molecules, constructs, or other objects with your colleagues,
or use them simultaneously on several computers (for instance, at work and at home).

In the Vector NTI archives:

- All DNA/RNA molecule information is written to and read from an archive file. This 
  includes molecule component fragments (if the molecule is constructed from other mole-
cules) and parent-descendant connections between molecules.
- Vector NTI automatically checks the consistency of molecule archive information adding 
  necessary parents (including DNA parents of translated protein molecules) or discon-
necting them if you have neglected to transfer them to the archives.
- When the archive is loaded into a new database, Vector NTI checks the information con-
sistency on any of database molecules and recalculates them if necessary.

**Opening Vector NTI Database Explorer**

The Vector NTI Database Explorer can be opened from several sources:

- Vector NTI Database Explorer can be opened and used independently by selecting Vector NTI Explorer in the Vector NTI Advance options under Start > Programs on your 
  computer. Select Start > Programs > Invitrogen > Vector NTI Suite 10 > Vector NTI Explorer.
- From a Molecule Display window, select File > Local Database or click the Local Data-
  base button ( ) to open Database Explorer.
- The Vector NTI Database Explorer can also be configured to open automatically in the 
  Vector NTI workspace each time a Vector NTI session is initiated. From a Molecule Dis-
play window, select Edit > Options from the menu bar. On the General tab of the 
  Options dialog box, select Open Local Explorer At Startup.

In this User’s Manual, the term “Explorer” refers to Vector NTI Explorer, not to Windows 
Explorer.
Elements of Vector NTI Database Explorer

The Vector NTI Database Explorer window consists of a menu bar, a toolbar, Table drop-down menu in the upper left corner, a Subset Pane on the left and a Database <Object > Pane on the right of the window (Figure 17.1).

Database Tables

The Vector NTI Database consists of the following eight tables:

- **DNA/RNA Molecule Table** – Stores DNA and RNA Molecules, including constructed and designed molecules. Three types of DNA/RNA molecules can exist in the Local Vector NTI Database Explorer:
  - Circular ( ) DNA Molecules
  - Linear ( ) DNA/RNA Molecules
  - Invitrogen ( ) Proprietary Molecules

  **Note:** Among the different types of molecules present in the Vector NTI Database are Invitrogen™ proprietary molecules. Over one hundred of the most commonly used, fully annotated, Invitrogen vectors are available in the database. When working with these specific molecules, included in a Molecule Viewer Text Pane is a hyperlink to the Invitrogen website providing you with relevant information about reagents specific to the annotated features of that molecule.

- **Protein Molecules Table** – Stores Protein Molecules.
- **Enzymes Table** – Stores restriction enzymes.
- **Oligos Table** – Stores oligonucleotides up to 1000 bases in length.
- **Gel Markers Table** – Stores gel markers that can be used in Vector NTI’s gel simulation function.
- **Citations Table** – Stores citations retrieved through the Vector NTI PubMed Search engine, from the NCBI website or from other searches. In addition, the Citation Table section of Database Explorer provides complete bibliographic conversion features described in the **Citation Table** section below.
- **BLAST Results Table** – Stores BLAST results retrieved using the BLAST Search module. BLAST Results can be opened in the BLAST Viewer module and saved to the Vector NTI Database from there.

- **Analysis Results Table** – Stores various analyses performed in Vector NTI and Analysis Monitor (see page 559), such as PCR analysis, PFAM, etc. When a saved analysis is opened from the Vector NTI Database, the associated molecule opens with the analysis information displayed in the Text, Graphics and Sequence Panes. (The Analyses display option must be turned on. See page 221).

The Vector NTI Database Explorer window displays the contents of one table at a time. You can alternate between tables by selecting a table from the Table menu or the Table drop-down list in the top left corner of the Explorer window.

**Citation Table**

One of the most convenient features of the Citation Table is its ability to generate citations and bibliographies for theses, dissertations and other publications. Bibliographies are generated according to the rule of scientific journals.

**Formatting Bibliographic References**

To format each bibliography record, you must insert a tag in the document where the reference is to be cited. In Database Explorer Citation Table, select the desired reference. From its associated shortcut menu, select the **Copy Tag** command. Then go to your word processor document and paste it. For example, the tag at this point might read [Smith, J. 1994 #213].

After finishing the writing, save your document as Rich Text Format (*.rtf) file.

When you are ready to prepare your manuscript, select the Citations Table in Database Explorer. Open a shortcut menu in the subsets pane and select **Format Manuscript**. In the dialog box that opens (Figure 17.2), browse for the RTF file you want to format.

In the Format Reference For drop-down menu, select a journal name. Click the **Start** button. The tagged document will be formatted into the document with properly formatted citations and bibliographies.

![Figure 17.2 Format Manuscript dialog box](image)

To place selected citations onto the Clipboard formatted according to the rules of specific journal, select the citation(s) and open the shortcut menu in the Database Objects pane. Select
Copy Bibliography. In the dialog box that opens (Figure 17.3), select the journal whose formatting the bibliography should follow. Click OK.

![Copy Bibliography dialog box](image)

Then proceed to your word processing document and paste the citations. They will be in the format for the journal you selected.

**Note:** In the Database Explorer, one cannot edit citations downloaded from NCBI.

### Database Subsets

Each database table contains a list of subsets, groups of objects (molecules, enzymes, etc.) organized into logical categories. You might have one subset for each of your molecule families, one for each taxonomic group, etc.

Subset separators are used to group subsets into logical categories by providing a visual separator or division between subsets in the Subset Pane. For example, you may create separators to divide subsets containing molecules used for different projects. You can add separators by right-clicking in the Subset Pane and choosing **Add Separator** from the shortcut menu. Once you have created a separator, you can change its position in the Subset Pane by clicking and dragging it to a new location in the Subset Pane.

- The principal subset is called Main and all database objects are always included in Main. Other subsets are merely windows into Main, showing a subset of the contents of Main; they do not serve as containers for molecules. Because of this, a database object can be included in any number of subsets without increasing the amount of disk storage the object occupies.
- Subsets are created, edited and deleted through the Database Explorer.
- When you select a subset in the left pane, Explorer lists the subset contents in the right pane and displays table- and subset-related commands on the menu bar and toolbar buttons on the toolbar. When you select one or more database objects in the right pane (activating the Object Pane), the menu bar displays object-related menu options and toolbar buttons.
- To alternate between panes, click on a pane to activate it or use the F6 keyboard shortcut. To open context-sensitive shortcut menus, click on subsets or database objects with a right click. The shortcut menu displays commands appropriate for the object under the cursor.

### Database Objects

Each database object is described by a set of fields, describing various object data. Each object type defines its own set of fields. System fields are common for all database objects; other fields are specific for a particular object type. A user may also define additional user fields.
Database Object System Fields

Most of the object system fields, assigned names by Vector NTI, are accessible from various parts of the program’s user interface. Some of them, however, are only visible in specialized object type-specific dialog boxes or display windows.

The following is the full list of named system fields displayed by Vector NTI:

<table>
<thead>
<tr>
<th>Field name</th>
<th>Description</th>
<th>Used in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name of an object</td>
<td>NPEO, OC</td>
</tr>
<tr>
<td>Local/Foreign</td>
<td>Object’s ‘ownership’ status in the local database</td>
<td>NPEO</td>
</tr>
<tr>
<td>Original Author</td>
<td>Original author (creator) of an object</td>
<td>NPEO</td>
</tr>
<tr>
<td>Created</td>
<td>Date of an object’s creation (See Date and Time section following)</td>
<td>NPEO</td>
</tr>
<tr>
<td>Author</td>
<td>Author of the last modification to an object (see following section)</td>
<td>NPEO</td>
</tr>
<tr>
<td>Modified</td>
<td>Date of the last modification to an object</td>
<td>NPEO</td>
</tr>
<tr>
<td>Submitted by</td>
<td>The name of a Vector NTI user who submitted the object</td>
<td>NPEO</td>
</tr>
<tr>
<td>Description</td>
<td>Description string</td>
<td>NPEO</td>
</tr>
<tr>
<td>Form</td>
<td>Circular or Linear</td>
<td>N</td>
</tr>
<tr>
<td>Storage Type</td>
<td>Basic, Designed or Constructed</td>
<td>N</td>
</tr>
<tr>
<td>Fully Processed?</td>
<td>Molecule’s construction status</td>
<td>N</td>
</tr>
<tr>
<td>Length</td>
<td>Length of a molecule/oligo/enzyme’s recognition string in bp</td>
<td>NPOE</td>
</tr>
<tr>
<td>DNA/RNA</td>
<td>Nucleic acid type</td>
<td>NO</td>
</tr>
<tr>
<td>Degenerate?</td>
<td>Is an oligo degenerate?</td>
<td>O</td>
</tr>
<tr>
<td>Oligo Sequence</td>
<td>Oligo’s nucleotide sequence</td>
<td>O</td>
</tr>
<tr>
<td>Recognition String</td>
<td>Enzyme’s recognition string</td>
<td>E</td>
</tr>
<tr>
<td>Terminus Type</td>
<td>Enzyme’s terminus type</td>
<td>E</td>
</tr>
<tr>
<td>Palindromic?</td>
<td>Is an enzyme palindromic?</td>
<td>E</td>
</tr>
<tr>
<td>Ambiguous?</td>
<td>Is an enzyme ambiguous?</td>
<td>E</td>
</tr>
<tr>
<td>Number of Fragments</td>
<td>Number of fragments in a gel marker</td>
<td>G</td>
</tr>
<tr>
<td>Longest Fragment</td>
<td>Length of the longest fragment in a gel marker</td>
<td>G</td>
</tr>
</tbody>
</table>

Table 17.1 Named system fields displayed by Vector NTI
In addition to the system fields named above, every database object has comments and key-
words associated with it. These fields are accessible from within specialized object editors and
Display windows.

Note: All automatically updated fields, user-defined fields and comments are searchable. Using the
database search interface described later in this chapter, you can perform string context
searches through any of the fields above for any database objects.

Of the general system fields data itemized in the above table, only Author and Date and Time
display features are independent of object types. Their sources and modification options are
summarized below.

**Author Information**

Information regarding authors and times of creation and modification for all database objects is
stored in the following named System Fields:

- **Original Author**: name of Vector NTI user who created the object,
- **Author**: name of Vector NTI user who last modified the object,
- **Creation Date**: date and time when the object was created,
- **Modification Date**: date and time when the object was last modified in the local Vector
  NTI database.

When you start Vector NTI for the first time, in the Author Information dialog box you can identify
yourself as Vector NTI user and author of all modifications and new objects you create in the
local database. By default, it presents the name given to the Vector NTI Installation program.
This information is automatically attached to all database objects you create and/or modify when
you export them, copy to shared databases etc.

Note: Changes you might make later to Author Information do not affect information stored in data-
base objects you submitted to shared databases earlier.

<table>
<thead>
<tr>
<th>Field name</th>
<th>Description</th>
<th>Used in:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortest Fragment</td>
<td>Length of the shortest fragment in a gel marker</td>
<td>G</td>
</tr>
<tr>
<td>Telephone</td>
<td>Phone part of the Contact info</td>
<td>C</td>
</tr>
<tr>
<td>E-mail</td>
<td>E-mail part of the Contact info</td>
<td>C</td>
</tr>
<tr>
<td>Field Type</td>
<td>User field type</td>
<td>U</td>
</tr>
<tr>
<td>Indexed?</td>
<td>Is user field cached/indexed?</td>
<td>U</td>
</tr>
<tr>
<td>Default</td>
<td>Default value of a user field</td>
<td>U</td>
</tr>
<tr>
<td>List of Pre-defined Values</td>
<td>List of predefined values of a user field</td>
<td>U</td>
</tr>
<tr>
<td>Record #</td>
<td>Database record ID number</td>
<td>NPEOG</td>
</tr>
</tbody>
</table>

Table 17.1 Named system fields displayed by Vector NTI (continued)
objects while you view or edit them, the data does follow the object and is displayed when exported or viewed by another VNTI user.

- When you import an object or transfer it from a VNTI Shared Database into your local database, if you are the author the object is still marked as Local. If the author is someone else, then the object is marked as "Foreign" and the author information is always displayed when you are viewing or editing the object.

- The original author (creator) of an object, either "Local" or "Foreign", never changes. If you modify a "Foreign" object, however, it becomes "Local" and you become its new "immediate" author.

- If you are importing an object in which author or original author information is not stored in your contact database, VNTI lets you update the contact and author data in your database.

To review or edit author data, select **View > Author Info…** opening the Author Information dialog box where you can do that.

**Date/Time Display**

To change the date/time display format, select **Edit > Options** on the menu bar of the main Vector NTI’s window (Molecule Display Window). On the General tab of the Options dialog box, select the standard US date format (for example, 03/20/96 04:20PM) or a regional date/time format, the default format recognized by your computer. Change this format using the Windows Control Panel/Regional (or International) dialog box. (For more information regarding formatting, see Chapter 3.)

Authors, original authors, and times of creation and modification for all database objects are updated automatically by Vector NTI and should not, in normal circumstances, be modified manually. If such a need occurs, use the **<database object > > Set System Attributes** command for setting values of these fields in the Vector NTI Database Explorer.

**User-defined Fields**

For information management functions, user-defined fields are associated with any database object (molecules, oligos, etc.). A User Field containing laboratory-specific and user-specific descriptions can be imported, exported, displayed, printed and managed just as the regular database fields.

While previously defined User Fields are maintained and presented when you create new database objects, imported objects may contain fields not described on the local list. Vector NTI helps you to add their descriptions to the list whenever you try to edit the information contained in these fields. Vector NTI also allows you to import and export User Field Definitions across several VNTI databases.

The "User Field Manager" section on page 197 of this chapter contains the detailed description of all User Fields-related operations. For a detailed example of how to use the User Fields Manager, walk through the Tutorial: Database Explorer in Chapter 7.

**Comments**

Each database object can contain a text comment of unlimited length. Vector NTI automatically converts stored comments when importing or exporting molecules from/to GenBank/GenPept and EMBL/SWISS-PROT files.

Comments can be edited in the Comments tabs of the objects’ editors or immediately in the Molecule Display window (for molecules only). In a Molecule Display window, double click on the Comment line or click on the Comment icon to open the Comment Editor containing the molecule comment.
Keywords

Each database object can contain a list of associated keywords that are very useful for database searches. Keywords can be edited in the Keywords tabs of the object’s editors.

To enter keywords, in the keywords text box, type a new word or select an item from the list of existing keywords. Click the <Add button to move the keyword into the keyword list. To remove an item from the list of keywords, select the item(s) you want to remove and click the >Remove button.

Database Search

You can perform a database search to locate any type of object stored in the database. In general, search procedures launched from Database Explorer follow these basic steps:

1. Select the appropriate Database Table type and select Database > Search on Explorer’s menu bar or click the Search button ( ) on the toolbar.
2. In the <Object > Database Search dialog box, verify the Database Table type.
3. In the Search dialog box, (Figure 17.4) check the boxes for the filters to be included in the search. Press the enable filter buttons to open dialog boxes.
4. In the Conditions Filter dialog boxes that open, define additional search parameters. (See detailed descriptions of filters below.) Only those molecules that satisfy all the conditions you define are found in the search. Click OK in each to return to the Search dialog box.
5. When you are satisfied with the search setup, click the Start Search button.
6. If Vector NTI finds any objects by a database search, the appropriate Subset Name dialog box appears. Enter a name of new subset or choose a subset name from those listed. All objects found will be placed in the specified subset overwriting the existing subset content (if any).

Notes:

- If the specified subset already exists, Vector NTI emptied it before adding found objects. You can reﬁne your search by using the resulting subset to limit the search space of the next search.
- If you check multiple Filters in the Molecule Database Search dialog box, a search that satisfies all the checked criteria (an AND search) is performed. If you want to search according to only one filter criterion, remember to leave all other filters, except the one of interest, unchecked.

Database search filter parameters specific to particular database objects are outlined below. Specific settings are not discussed unless they need clarisation.
DNA/RNA and Protein Molecule Database Searches

The Molecule Database Search dialog box (Figure 17.4) for DNA/RNA and protein molecules are identical except for the database shown in the Look In drop-down menu. Differences in the specific filter setup dialog boxes are specified in the discussion for each.

![Molecule Database Search dialog box](Figure 17.4 Molecule Database Search dialog box)

To set the conditions for a particular search filter, check the appropriate filter checkbox and press the corresponding filter button. For example, if you want to set the parameters for a Text Search, check the **Text Filter** checkbox and then click the **Text Search Setup** button.

**Searching by Object Name in the Vector NTI Database**

**Database Search Name Filter**

To perform a Name search, check the **Name** checkbox and type the name you want to search for in the textbox in the Name section of the Molecule Database Search dialog box. If you want the search to match the case of the entered name, check the **Match Case** checkbox.

The wildcard characters * and ? are allowed in a name search and have the following significance:

- * = any number of characters (including no characters)
- ? = any single character

**Alternate Name Search Method**

You can perform a quick name search in the Database Exploring window by performing the following steps:

1. Open the Exploring – Local Vector NTI Database window.
2. Navigate to the database table of interest by choosing it from the Database Table drop-down list.
3. Sort the objects alphabetically according to name (if not already sorted by name) by clicking on the Name column heading.
4. Highlight the first name on the list.
5. Using the keyboard, type the first letter (or first few letters, quickly) of the desired object name. The list scrolls to that area on the list.
Attributes Search Conditions Box (DNA)

![Attributes Search Conditions Box (DNA)](image)

If all conditions in a group are checked, or if all conditions in a group are blank, that group has no effect on the search.

Attributes Search Conditions Dialog Box (Protein)

![Attributes Search Conditions Dialog Box (Protein)](image)

Text Search Conditions Dialog Box (DNA and Protein)

![Text Search Conditions Dialog Box](image)

To add a new substring, enter the text and click the < Add button (Figure 17.7). To remove a substring, highlight it and click the > Remove button. Check the fields to be searched for the text substrings. Check Match Case for case-sensitive text substrings.
Keywords Search Conditions Dialog Box (DNA and Protein)

![Keywords Setup dialog box](image)

Figure 17.8 Keywords Setup dialog box

Click the < Add button to select from the list of existing keywords (Figure 17.8). To remove a keyword from the box, select it and click the > Remove button. To clear the Keywords box, click the >> Remove All button.

**Note:** You cannot add keywords in this dialog box.

Ancestors Search Conditions Dialog Box (DNA and Protein)

![Ancestors Setup dialog box](image)

Figure 17.9 Ancestors Setup dialog box

Molecules that have as a direct ancestor any one of the molecules shown in this dialog box will be found by the search. *Ancestors for proteins in this context are DNA molecules.*

To add a molecule to this box, click < Add. Select from the database molecules list and click OK. To remove a molecule from the ancestors list, select the molecule in the dialog box and click > Remove, or to clear the list, click >> Remove All.
Oligo & Peptides Search Conditions Dialog Box (DNA and Protein)

Click the Add Oligos button to add an oligonucleotide for similarity search. In the Oligonucleotides dialog box that opens, you can select the oligo(s) from the Vector NTI Database or create new oligo(s) for the search.

To add a peptide for similarity search, click the Add Peptide button. In the Peptide dialog box that opens, enter the peptide name and query sequence, and description, if desired.

To remove a search condition from the list box, select the condition and click Delete. To clear the Oligonucleotides/Peptides box, click Clear.

Select the genetic code you want to use for the search from the Genetic Code drop-down box.

Feature Search Conditions Dialog Box (DNA and Protein)

To add a feature to the search conditions, click the < Add button (Figure 17.11). In the dialog box that opens (Figure 17.12), define restrictions for the molecule search. Feature types differ for DNA/RNA searches and protein searches.
Select a feature type in the left pane and optionally, using the radio button, one or more feature names. If you define a type of feature but no specific feature names, the search condition is defined as any feature of that type.

To remove a search condition from the Features list, select the condition and click > Remove. To clear the Features list, click >> Remove All.

Enzyme Database Search

When you launch an enzyme search, the following Search dialog box opens (Figure 17.13):

- **Recognition String Filter**: Enter a nucleotide string that must be found in the recognition site of the desired enzyme.
- **Attributes Filter**: Specify enzyme attributes necessary for the search. “Ambiguous” recognition site means the site containing any nucleotide besides A, T, G, or C.
- **Text Filter**: Check the Text Filter button to filter search results based on text inside various object fields. Refer to page 173 for a description of the Text Filter Condition.
- **Keywords Filter**: Refer to page 174 for a description of the Keywords filter. **Note**: You cannot add new keywords in this dialog box.

Oligo Database Search

When you launch an Oligo database search, the following dialog box opens (Figure 17.14):
- **Nucleotide String Filter**: Enter a Nucleotide String filter that must be found in the sequence of the desired enzyme and select the strand to search, direct or complementary.

- **Attributes Filter**: Check to limit the search space to DNAs only or RNAs only. If both are checked or unchecked, the filter has no effect.

- **Text Filter**: Check the **Text Filter** button to filter search results based on text inside various object fields. Refer to page 173 for a description of the Text Filter Condition.

- **Keywords Filter**: Refer to page 174 for a description of the Keywords filter. **Note**: You cannot add keywords in this dialog box.

**Gel Marker Database Search**

When you launch a search in the Gel Marker Database, the following dialog box opens (Figure 17.15):

![Gel Marker Database Search dialog box](image)

- **Attributes Filter**: Check to limit the search space to gel markers with specified length of the longest fragment. If all are checked or unchecked, the filter has no effect.

- **Text Filter**: Check the **Text Filter** button to filter search results based on text inside various object fields. Refer to page 173 for a description of the Text Filter Condition.

- **Keywords Filter**: Refer to page 174 for a description of the Keywords filter. **Note**: You cannot add keywords in this dialog box.
Citation Database Search

When a Citations Database Search is launched, the following dialog box appears, allowing you to search by Citation, Text and/or Keywords:

![Citations Database Search dialog box](image)

**Figure 17.16 Citations Database Search dialog box**

BLAST Results and Analysis Results Database Searches

When a BLAST Results or Analysis Results Database Search is launched, the following dialog box appears, allowing you to search by Attributes, Text and/or Keywords:

![BLAST Database Search dialog box](image)

**Figure 17.17 BLAST Database Search dialog box**

Database Management Operations

Backup, Restore and Cleanup

Three general functions (all initiated by choosing options under the Database menu of Database Explorer) relate to the entire database:
- **Database Backup** makes a copy of your database in a specified directory or folder. For instance, you may back up the data to removable storage to make a permanent, safe record of your database. Making such a backup is the only way to be sure your data is safe in the case of hard drive failure.

- **Database Restore** restores all database files from the last backup. Your database is restored from a specified directory or folder, overwriting all current database files.

- **Database Cleanup** removes from the data files and database tables all the “garbage” data that may accumulate when you extensively create and delete database objects. This also serves as a recover function when files or tables are lost or damaged. Although these data do not affect Vector NTI functionality, it is wise to perform Database Cleanup once a week.

**Important:** Vector NTI Database Backups are NOT cross-platform compatible (i.e. the Vector NTI Database on a PC cannot be restored using a VNTI Database Backup made from Vector NTI on a Macintosh, and vice versa. If you want to move the VNTI Database from PC to Macintosh (or vice versa), you must do so using VNTI Archive files. For more information on importing and exporting VNTI Database objects as Archives, see page 181 and page 184, respectively.

**Selecting the Local Database**

The Select Local Database command in Database Explorer allows you to direct Vector NTI to the local Vector NTI database of interest, in the case where you have created multiple local databases. For example, you may create additional Vector NTI databases by copying the existing local database to a different location on your hard drive. In this case, you can use the Select Local Database command to alternate Vector NTI’s focus to the different local databases.

**Note:** Vector NTI/Database Explorer can only view one database at a time.

To point Database Explorer to a different Vector NTI local database, choose **Database > Select Local Database** from the Database Explorer menu, navigate to the Vector NTI database of interest in the Choose Local Database Directory dialog box that opens, highlight the database folder and click the **OK** button. The selected Vector NTI database now appears in the Database Exploring window.

**Note:** You must have all other Vector NTI Advance programs closed, except Database Explorer, when executing the Select Local Database command.

**Operations on Subsets**

To initiate any of the following operations on a subset, select the subset name in the left pane of Database Explorer. Press the toolbar buttons as described or choose Table or Edit or corresponding shortcut menu commands. Descriptions of each operation are included below.

**Creating a New Subset**

To create a new subset and/or add objects to it, perform one of the following options:

- Select **Table > New > Subset** or click the **New Subset** button ( ) to create a new (empty) subset with a default name, Group 1. You can edit this name in the enabled text box; press ENTER to confirm.

- Select **Table > New > Subset (Using Subset Editor)** to create a new subset and select its contents using the Edit Subset dialog box described below under Editing a Subset.

- A simple way to add new molecule or other object to a subset is to open the source subset in the Subset Pane. Select and drag selected objects from the list to the new subset. The objects are now listed in the original subset and the new subset.
To create a new subset and place objects in it at the same time, select objects for the new subset in the Database Objects Pane and click the New Subset button ( ).

### Editing a Subset

In the Subset Pane, select a subset to be edited and select Table > Edit Subset. This option is not available for subsets where the content is maintained automatically. This opens the Edit Subset dialog box (Figure 17.18):

![Edit Subset dialog box](image)

- The Source Subsets column shows all of the subsets in your Vector NTI database table except the one you are editing. The buttons below this column control how subset objects in subsets selected in the Source Subsets column are displayed in the Source Content column:
  - **Intersection**: Each molecule that is present in all highlighted subsets is displayed in the Source Content column. Molecules not included in all selected subsets are thus excluded.
  - **Difference**: Each unique molecule from selected subsets, in other words, a molecule present in only one of the highlighted subsets is displayed. Molecules that occur in more than one subset are thus excluded.
  - **Union**: All molecules that are present in any of the selected subsets is displayed. If only one subset is highlighted, the Intersection, Difference, and Union buttons are disabled, and all the objects in that subset are shown in the Source Content column.

- The Source Content column shows combinations of objects present in the subsets highlighted in the Source Subsets column. See descriptions under the previous bullet.

- The Resulting Content column shows the contents of the subset you are editing.

To add objects to a subset, highlight one or more subsets in the Source Subsets Column. In the Source Content column, highlight the object names or use the Select All and Unselect All buttons in that column to highlight or deselect all objects at once. When the objects you want are highlighted, click the Add >> button to add them to the Resulting Content column. Click OK to exit the dialog box. The objects are now listed in the original subset and the new subset.

### Removing Molecules from a Subset

In the Subset Pane, select the subset to be edited. Select Table > Edit Subset or Edit from the shortcut menu. The Resulting Content column in the Edit Subset dialog box (Figure 17.18) shows the current contents of the subset you are editing. To remove objects from the subset,
highlight their names in the Resulting Content column and click the << Remove button. Select All and Unselect All buttons below the Resulting Content column allow you to highlight or deselect all objects in the column. Click OK to exit the dialog box. Database objects that are removed continue to exist in the database and other subsets where they were included before this operation.

For alternative ways of removing objects from a subset, see Operations on Database Objects on page 185.

**Rename Subset**

Select Table > Rename Subset or click the Rename button ( ). In the enabled text box, edit the name. This operation does not affect the subset contents or database objects. This option is not available for subsets where the content is maintained automatically.

**Dismiss Subset**

Select Table > Dismiss Subset or click the Dismiss Subset button ( ). This operation removes a selected subset, but does not affect the database objects. They are still stored in the database and exist in other subsets where they were included before this operation. To empty the subset first, use the Clear Subset operation.

**Clear Subset**

Select Table > Clear Subset to exclude all objects from the selected subset. The subset itself is emptied but not removed. This operation does not affect database objects; they are still stored in the database and included in all other subsets where they were included before this operation. This option is not available for subsets where the content is maintained automatically.

**Delete Contents**

Select Table > Delete Contents from Database to physically delete from the database all objects of the specified subset. If you choose OK in the confirmation dialog box, the objects in the subset are permanently deleted from the Vector NTI database, and the subset itself is removed.

**Subset Summary**

Select Table > Subset Summary to create a text file containing a summary of the specified subset. In the Write Summary To dialog box that opens, indicate the directory or folder and file name where the summary will be stored. Click the Save button and Vector NTI creates the summary.

**Subset Properties**

Select Table > Properties or the Properties button ( ) to display properties for the selected subset.

**Import Molecules into a Subset**

To import molecules into a subset using drag and drop, use one of the following techniques:

- **Import Molecule(s) or Sequence(s) From Text File(s):** Highlight the molecule file name(s) in Windows Explorer (Note: molecule files must be in accepted formats, see below). Drag and drop the highlighted name(s) onto the desired subset in the VNTI Database Exploring window.
- **Import Directory of Files:** For a folder containing multiple molecule files (in accepted formats, see below), drag and drop the folder onto the desired subset in the Exploring
Local VNTI Database window. All the molecules in the folder are imported into the subset.

To import molecules into a subset using menu options, select Table > Import, then one of the following options:

- **Import Molecule From Text File** imports DNA/RNA or protein molecule data from a GenBank/GenPept, EMBL/SWISS-PROT, or FASTA file. Vector NTI imports all molecule data including feature table (if any).

  - In the Import Molecule dialog box, choose the appropriate import file format—GenBank, EMBL, or FASTA for DNA/RNA molecules and GenPept, SWISS-PROT or FASTA for proteins and click OK. In the Import dialog box, browse for the source and file to be imported and click Open. Vector NTI reads the file, checks all data, creates the new molecule, and displays its data in the Molecule Editor dialog box. After entering or editing the data, press OK. Vector NTI saves the new molecule in the database.

  - **Import Sequence From Text File** imports a nucleotide or amino acid sequence text file, creating a new DNA/RNA or protein molecule with the sequence. *This must be in ASCII format.* In Import dialog box, browse for the source and file to be imported and click Open. Vector NTI reads the file, checks all data, creates the new molecule, and displays its data in the Molecule Editor dialog box. After reviewing or editing the data, click OK, saving the new molecule in the database.

    - For ASCII format details, see Appendix C.

  - **Import Molecules From Archive** imports objects from a Vector NTI archive. In the Import dialog box, browse for the source and archive file to be imported and click Open.

  - **Import Molecules From Directory of Text Files,** (option for molecules and oligos only), imports molecules and oligonucleotides from text files that must be in the GenBank format (DNA/RNA molecules), GenPept format (protein molecules) or Vector NTI's oligo text format (oligonucleotides). In the Import dialog box, browse for the List File and open. *When Vector NTI exports the objects into a directory of text files (see below), it automatically creates the appropriate List File.*

    - Notes on Molecule Import.

      - In Vector NTI versions prior to 6.0, import and export functions could be performed from the Molecule Display window. In version 7.0, they can be performed only through Database Explorer.

      - GenBank, GenPept and EMBL feature tables usually do not contain names for features. When importing, Vector NTI assigns “artificial” names for the features so that they will be recognizable on a graphics map. You can change features’ names after import using Vector NTI's Molecule Editor, or assign names to features before import, by adding “\texttt{label=<name >}” descriptors in the flatfile for any desired features. Here is a sample GenBank/EMBL feature description with a defined feature name:

        CDS 86..1276
        /label=TC(R)

    - For more information about GenBank/GenPept /EMBL Feature Table descriptors, see GenBank/GenPept /EMBL documentation.

**Import Enzymes into a Subset**

Restriction endonucleases can be imported into Vector NTI from the REBASE database, which is available through anonymous ftp site, ftp.neb.com. To obtain further information, refer to the website: [http://rebase.neb.com/rebase/rebase.html](http://rebase.neb.com/rebase/rebase.html). *For a recent review of the REBASE database see Nucleic Acids Research 22: 3628-3639, 1994.*

Each REBASE release contains several data files in different formats. Vector NTI uses bairoch format, so you can download the bairoch.### file where ### indicates the current version (for
example, 506 indicates June 1995). Distributive version of Vector NTI contains the last release of the REBASE bairoch.### file in the Vector NTI working directory.

To import restriction endonucleases into the Enzymes Table, select **Table > Import**, then one of the following options:

- **Enzymes From Archive**: Browse for the Archive file and open.
- **From REBASE Database**: Browse for **Vector NTI Suite > bairoch.###file** (see above) and select it in the Read REBASE Enzymes dialog box (Figure 17.19). Click **Open**. In the dialog box that opens, select the subset or name a new subset where objects will be inserted. *This file includes all enzymes in the bairoch file.*

![Figure 17.19 Read REBASE Enzymes dialog box](image)

- **New Enzymes From REBASE Database**: Browse for **Vector NTI Suite > bairoch.###file** and select it in the Read REBASE Enzymes dialog box (Figure 17.19) and click **Open**. In the dialog box that opens, select the subset where objects will be inserted. *This file includes only enzymes in the bairoch file not currently included in the Vector NTI Enzymes database.*

Vector NTI reads the file and checks the data it contains. If Vector NTI detects any inconsistencies between Commercial Source designations/codes in the VNTI Restriction Enzyme Database and those designated in the bairoch file, the Commercial Sources Import dialog box appears (Figure 17.20).

**Note:** For a Commercial Sources codes key, see the REBASE text file.

![Figure 17.20 Commercial Sources Import dialog box](image)

Your are given the choice to 1) merge the Commercial Source codes in the imported file with the current codes in the Database, 2) overwrite the current codes with the incoming codes, or 3) decline importing Commercial Sources data. Choose the appropriate option and click the **OK** button.
Vector NTI then loads the REBASE enzymes in the file into the REBASE Enzymes dialog box (Figure 17.21):

![REBASE Enzymes dialog box](image)

Figure 17.21 REBASE Enzymes dialog box

To filter the enzymes in the Source Enzymes/Selected columns, check on or off the appropriate boxes in the Use REBASE Subset group in the lower right corner of the dialog box and click the Hide Others button. This is particularly convenient when you wish to identify restriction enzymes based on certain criteria.

When all the desired enzymes are in the Selected list, click the Import Selected button to import them into Vector NTI.

Export Subset

**Export Subset Into Vector NTI Archive**

Select Table > Export > Subset into Archive to export all objects of the specified subset to a VNTI archive. In the Write To dialog box, indicate the directory or folder, and file name for archive storage. Click the Save button.

**Export All Subset Objects Into Directory of Text Files (molecules and oligos only)**

Select Table > Export > Subset to Directory of Text Files to export all objects of the specified subset into a directory of text files in GenBank (DNA/RNA molecules), GenPept (protein molecules) or Vector NTI’s oligo text format (oligonucleotides). In the dialog box, browse for the location and enter the file name under which the subset is to be stored.

Copy and Save or Print a List of the Objects in a Subset

You can copy and save/print a list of the objects in a subset, including the object names and other desired column information by performing the following steps:

1. Open the Exploring Local Vector NTI Database window.
2. Navigate to the database table of interest by selecting it from the Database Table drop-down box.
3. Right-click on the column headings in the Database <Object> Pane (right-hand pane) and choose Columns... to launch the Columns dialog box.
4. Add or remove columns to customize the information showing in the Database <Object> Pane according to what you want to print/save.
5. Close the Columns dialog box.
6. Create a subset and add to it the objects you want to print/save or navigate to the desired subset, if already created. (Alternately, highlight the particular objects of interest, instead of creating a subset.)
7. With the Database <Object> Pane activated, click the **Camera** button on the toolbar or choose **Edit > Camera** from the menu to launch the Camera dialog box.

8. Select the desired Range (**Selection** or **All**) and **Clipboard** as the Copy To location.

9. Open Microsoft Word or Excel and paste the data from the Clipboard.

10. Print and/or save the data from Word/Excel.

**Operations on Database Objects**

To perform various operations on an individual database object or group of objects selected in the Object Pane of the Database Explorer, select menu options from the left-most menu option that correspond to the currently displayed table type. For example, when the **DNA/RNA Molecule** table is selected, the left-most menu option is **DNA/RNA Molecules**, when the **Citations** table is selected, the left-most menu option is **Citations**.

**Notes about object operations:**

Most of the following options appear with all database objects.

Most commands are also available from the objects’ shortcut menu, opened with a right click. When available from toolbar buttons, they are shown.

Menu commands specific for certain objects are discussed after the following bulleted list.

- **Open**: opens display windows for selected objects; available only for DNA/RNA and protein molecules.

- **Edit**: includes editing options for an object selected in the Database Objects Pane, such as user fields, comments, keywords, etc. *This command is enabled only with a single selected object.* You can also use the **Edit** button ( ).

- **New**: includes options for creating a new subset for selected object(s) or creating a new molecule or object. For specific details, refer to creating new <specific object> in the index.

- **Import…**: operations are described in the previous section.

- **Export**: operations are described in the previous section.

- **Exclude from Subset** Tree Pane History view excludes selected objects from the current subset. Objects are still stored in the database and are included in all other subsets where they were included before this operation.

- **Delete from Database**: Physically deletes a selected object from the database. (Confirmation is requested before deleting the object.) *If a DNA molecule is deleted, Vector NTI scans all its descendants and disconnects them from the deleted molecule.* You can also use the **Delete** button ( ).

- **Rename**: renames the selected object. If a DNA molecule is renamed, Vector NTI scans all its descendants and also changes molecule’s name in all component fragment descriptions of all the molecule’s descendants. You can also use the **Rename** button ( ).

- **Duplicate**: makes copies of the selected database objects. The replicas of selected objects are created in the database and included in the current subset. The replica of an object named ‘x’ will be called ‘Copy of x’. *Note: Copies of objects are not related to original objects. Thus, “copies” do not track further changes you may make to the originals.*

- **Properties**: displays the properties of the selected object(s). If one object is selected, all the named object fields with their values are displayed. Some object data (like sequence and comments) are not stored in named fields and are not displayed in the properties dialog box. For molecules, the property dialog box has two extra tabs: Parents and Descendants, displaying Parent and Descendant trees for the selected molecule. You can also use the **Properties** button ( ).
- **Set System Attributes:** Allows manual modification of the automatically updated fields of the selected objects - Author, Original Author and Creation Date in the Set System Attributes dialog box. If checked, the drop-down menu (edit box in case of Creation Date) below the check box is activated for setting the value of the appropriate field.

  The Author and the Original Author fields can be set only with the names of contacts already stored in the contact database. The Creation Date field must be set with a date and time in the USA format: "MM/DD/YY H:MIN AM (or PM)" where H is from 1 to 12. For more information, refer to page 169.

**DNA/RNA Molecule Operations**

These operations are applicable to DNA/RNA molecules:

- **Table > New > Molecule (Using Sequence Editor)** creates a new DNA/RNA molecule, entering the nucleotide sequence manually using Sequence Editor. The Molecule Editor dialog box is described in Chapter 19 (see page 237).

- **DNA/RNA > Gateway Cloning** is an Invitrogen™ technology that functions as a universal cloning system with rapid and efficient routes to protein expression, functional analysis and cloning/subcloning of DNA segments. This menu option launches a Vector NTI 9 feature that provides an interlocking set of computational tools for planning a Gateway cloning experiment from beginning to end. (This option is also available from the shortcut menu opened with a right click when a DNA or RNA molecule is selected in the Database Objects Pane. For more information, see GATEWAY and TOPO Cloning, Chapter 23).

- **DNA/RNA > Delete with Descendants from Database** physically deletes the selected DNA molecule together with its descendants (DNA/RNA and protein molecules) from the Database. Confirmation is requested before deletion.

**Protein Molecule Operations**

This operation is applicable to protein molecules:

- **Table > New > Molecule (Using Sequence Editor)** lets you create a new protein molecule, entering the amino acid sequence manually using Sequence Editor. The Molecule Editor dialog box is described in Chapter 19 (see page 237).

**Enzyme Operations**

These operations are applicable to RENs (restriction endonucleases) only.

- **Table > New > Enzyme** opens the Enzyme Editor tabbed dialog box where you can create a new database enzyme in the Enzyme Editor (Figure 17.22):
Enter information about the new REN as follows:

- **General Tab**: Enter the name of the REN in the Name text box.
- **Enzyme Tab**:
  - **Recognition String** – Enter the enzyme’s recognition string in the Recognition String text box
  - **Cleavage Point/Methylation Base** – In the Cleavage Point/Methylation Base field on the Enzyme tab, enter the number of the nucleotide immediately after the direct-strand cleavage point. The following figure demonstrates how cleavage points of palindromic sites are defined.
    
    Cleavage Point = 2

    ![AATATT TTATAA](image)

    Cleavage Point on Complementary Strand – If the REN has a non-palindromic recognition site, the Cleavage Point on Complementary Strand field will be enabled. In this field, enter the number of the nucleotide immediately after the complementary-strand cleavage point. The following figure demonstrates how cleavage points are defined for non-palindromic sites on both direct and complementary strands:
    
    Cleavage Point on Complementary Strand = 10

    ![AAGTNNNNNN TTCANNNNNN](image)

    Cleavage Point on Complementary Strand = -4

    ![NNNNNNNAAGT NNNNNNTTCA](image)

    The preview box on the Enzyme tab helps you to enter cleavage points.
  - **Restriction Enzyme/Methylase** – Select the enzyme type, regular REN or Methylase. Note: The enzyme type can be displayed in a field in the Database (Figure 17.23) by adding the Enzyme Type column to the Enzyme Database columns display
(for more information on adding columns to the Database Explorer display, see Options dialog box Details View Tab on page 192):

![Figure 17.23 Enzyme Type field added to Enzyme Database display](image)

- **Type of Methylation** – If Methylase is selected for the enzyme type, the Type of Methylation field becomes enabled. Enter the methylation type.

**Note:** The enzyme type and methylation type, if applicable are indicated in the Restriction/Methylation Map folder in the Text Pane of the Molecule Viewing window for displayed cutters (Figure 17.24):

![Figure 17.24 Restriction/Methylation Map showing methylation position and type](image)

- **Description** – In the Description box on the Enzyme tab, enter the REN’s description.
  - **User Fields tab:** Add custom data in a form of fields. For details, refer to page 200.
  - **Comments tab:** Enter text comments.
  - **Keywords tab:** Enter key words for the REN, following the directions on page 171.

When all data is correct, choose **OK** to save the enzyme and return to the Database Explorer.

**Oligo Operations**

These operations are applicable to oligonucleotides only.
Table > New > Oligo opens the New Oligo dialog box where you can create a new database oligo (Figure 17.25). The New Oligo and Edit Oligo dialog boxes are the same.

![New Oligo dialog box](image)

Figure 17.25 New Oligo dialog box

Enter information about the new (or to-be-edited) oligo as follows:

- **General Tab**: Enter the name of the oligo.
- **Oligo Tab**:
  - Nucleotide Sequence: Enter or edit the oligo’s nucleotide sequence. Valid characters: ATUCG.
  - Select the oligo type.
  - Complementary box: Check to replace the oligo sequence with the complementary one.
  - Description box: Enter or edit the oligo’s description.
- **User Fields Tab**: Add custom data in a form of fields. For details, refer to page 200.
- **Comments Tab**: Enter text comments.
- **Keywords Tab**: Enter keywords for the oligo. For details, see page 171.

When all data is correct, choose OK to save the oligo and return to the Database Explorer.

**Analyze Oligo**: With an oligo highlighted in the Database Objects pane, select Analyses > Oligo Analyses > Thermodynamic Properties or Oligo Duplexes to open corresponding dialog boxes. These dialog boxes are both described in Chapter 21.

**Gel Marker Operations**

These operations are applicable to gel markers only.
Table > New > Gel Marker opens the Gel Marker Editor tabbed dialog box (Figure 17.26) where you can enter (or edit data) for a gel marker.

![Figure 17.26 Gel Marker Editor tabbed dialog box](image)

You may enter information about the new (or edited) gel marker as follows:

- **General Tab**: Enter the name of the gel marker.
- **Gel Marker Tab**:
  - Fragments list box: Lists by length all the fragments making up the marker. To add a fragment to the marker, enter its length in bp in the New Fragment box and click Add. The fragment is added to the fragment list. To remove a fragment from the marker, select it in the Fragments box and click the Delete button.
  - Description box: Enter the gel marker’s description.
- **User Fields Tab**: Add custom data in a form of fields. For details, refer to page 200.
- **Comments Tab**: Enter text comments.
- **Keywords Tab**: Enter keywords for the gel marker. For details, refer to page 171.

When the information is correct, choose OK to save the Gel Marker and return to the Database Explorer.

**Citation Operations**

The New/Edit Properties dialog box for Citation objects has the following standard object tabs: General, User Fields, Comments and Keywords. For information about these tabs, refer to the above sections.

The Edit Properties dialog box for Citation objects has the following unique tabs:

- **Citation Tab** – Contains information about the citation, such as title, authors and reference lookup information.
- **Abstract Tab** – Contains the abstract for the citation.
- **MeSH Tab** – MeSH is the National Library of Medicine’s controlled vocabulary thesaurus. The MeSH entries on this tab are terms or subject headings found in the selected citation. These terms can be reviewed, but not edited.

**BLAST and Analysis Results Operations**

The Edit Properties dialog boxes for BLAST and Analysis Results objects have the following standard object tabs: General, User Fields, Comments and Keywords. For information about these tabs, refer to the above sections.
Drag and Drop

You can transfer objects by selecting them and using the LEFT-CLICK + DRAG technique to move them to another location. The result of drag/drop is equivalent to copy/paste operations, except that drag/drop leaves the Clipboard contents unchanged.

If you RIGHT-CLICK + DRAG, when you drop, you have the option to Copy [Here] or Cancel the operation. In case of transfer between subsets of the same database, a third option, Move Here, allows you to combine two operations: add selected objects to another subset and exclude them from the current subset. The Move operation is not available if the source subset has its contents maintained automatically.

Edit Menu Commands

The menu options under Edit are available for all database objects and are typical of other Windows Explorer applications. They are described as follows:

Copy/Paste

- Copy copies selected database objects and places them in subsets of the same database, local and shared databases, and databases and the file system. Selected objects, their source and names are copied to the Clipboard. You can also use the Copy button ( ). This mode of Copy copies meaningful biological data to be pasted into other biological analysis/management applications.

Notes about Copy:

- Copy using the Camera button ( ), copies data to the Clipboard as it appears on the computer screen. This is useful for pasting into a word processor during the preparation of publication documents, etc.
- Because only one copy can exist on the Clipboard at one time, the Copy command replaces previous Clipboard contents.
- Paste pastes copied data in one of four possible locations: (You can also use the Paste button ( ).
- The same subset of the same database where they were copied. This is an alternative way to duplicate copied objects.
- Into another subset of the same database where they were copied.
- Into another database such as a shared database. The resulting actions create a temporary archive for selected objects1 and import it into the target database.
- To a file system or vice versa. If you paste copied objects to the Windows 95/98 or Windows NT Explorer, Vector NTI Explorer automatically creates an archive file containing copied objects2. If you paste files from the Windows Explorer to Vector NTI Explorer, the files are imported into the target database. Vector NTI Explorer is usually able to recognize the types of the pasted files automatically and take appropriate actions. When it cannot determine the type of imported files, it allows you to select the type or cancel the operation.

1. In case the transfer of molecules, the archive may also contain parents of the selected molecules. See the Explorer Options section for details.

2. For molecules, the type of the created file depends on the Explorer settings. See the Explorer Options section for details.
Note: If database objects are renamed or modified AFTER you have copied them to the Clipboard, when you paste the object(s), the changes subsequent to copying are not reflected. Remember that the actual data is collected when you copy it, not when you paste it.

Explorer Display Options

The Explorer's database object pane can work in one of four view modes, selected under the View menu or by clicking on the View Mode button arrow, displayed at the right end of the tool-bar (Figure 17.27): Icons, Small Icons, List and Details. By default, the Database Explorer shows objects in the Details mode.

Figure 17.27 Display options

In the first three modes, only object icons and names are displayed. The Details view displays object icons, names and the values of object fields displayed in columns.

Options Dialog Box

To view or modify options related to the current database table display in Vector NTI Database Explorer, select View > Options to open the Options dialog box (Figure 17.28). To set options for a table other than the active one, switch to that table first. All of your settings are preserved between Vector NTI sessions.

Figure 17.28 VNTI DB Explorer Options dialog box

Details View Tab

On the Details View tab, you can select the fields to be displayed (in columns) for each object type, change the order of columns and the text alignment of each column.

Currently displayed columns are listed in the Displayed Fields list box. To customize the displayed columns, in the Available Fields list box, select fields to be displayed and click the ( ) button. To remove fields, select them in the Displayed Fields list, and click the ( ) button. To modify the alignment of displayed data, select among the Alignment radio buttons. The
descending order in the list is the left to right display order of the columns. To modify that, select fields and use the Move Up and Move Down buttons. Click OK to apply the settings.

This page can also be opened with a right click on any column heading in Database Explorer. When opened from that source, it is called the Columns page.

**Status Bar Tab**

This tab allows you to choose the type of information displayed on the Status Bar for a single selected object in the object pane. The default selection, Description, describes the item selected. If you select one of the fields, that value for the selected object is displayed. The 1 item selected option simply describes the selection.

**Tips Tab**

When you pause the mouse pointer over an object in the list pane of the Explorer, a popup message (tip) appears containing information about the object under the mouse pointer. On the Tips Tab, you can configure what is displayed.

**Archives Tab**

This tab allows you to choose what should happen to parent molecules when they are not included in a list of selected molecules to copy from the database to another location. In the Archives tab, you can choose to expand the set of molecules to include all missing selected molecules’ parents, copy each molecule independently (losing all links to its parents), copy each molecule while maintaining parental lines via “open” links imported into a database, or copy sequences of selected molecules to a FASTA file.

**Copy Molecule(s) Tabs (DNA/RNA or Proteins Molecules Table)**

These two tabs let you select the preferred file format for a single molecule or a set of molecules export (either via copy/paste or drag/drop to the file system or other programs). Each tab includes detailed explanations of the available options.

**Note:** ‘Parents’ for proteins are ‘DNA parents’.

Click OK to apply your settings to database display. All of the settings in the Database Options dialog box are preserved between Vector NTI sessions.

**Modifying Database Object Column Widths**

Column width settings can be changed directly. Move the cursor to the divider between columns where it changes to a two-headed arrow. Drag the divider right or left with the cursor until the column is the desired width.

**Sorting Objects in Database Explorer**

If you are in the Details View Mode, click on the column header to sort column contents. Clicking on the same column header for a second time reverses the sorting order.

To select descending or ascending sorting order for columns, or sort database objects by name or any of the displayed fields in any view mode, select View > Arrange Icons.

**Database Reports**

Reports on Database Explorer’s contents, suitable for word processing and spreadsheet programs, can be generated. The Database Objects Pane should contain all objects you would like to include in the report. If you do not have a subset containing all objects you need, select them and create a new subset or generate the report on selected objects only.

To prepare a report:
1. Select a database table and object subset. Make sure Explorer is in the Details View mode.

2. To select the objects' fields to be displayed, right click on a column heading, select Columns, opening the Columns display dialog box. Move the fields between the Fields columns, as described above in the Display View Tab section. Click OK to apply the settings.

3. Sort your objects by clicking on column's headers.

4. Select Edit > Camera or press the Camera button ( ) on the Explorer's toolbar. In the Camera dialog box, select the objects to be included in the report its destination.

5. Switch to a word processor, spreadsheet or database application and paste/import the report data there.

6. Vector NTI copies the report data in one of two different formats, depending on the copy location:
   - Plain text format with the TAB character used as column separator and the newline character as a row separator (sometimes called the “CSV” format) when copying to the Clipboard.
   - Rich Text format (RTF) when copying to a file.

**Explorer Shortcuts**

Database Explorer offers many shortcuts to make your work more productive:

- Drag one or more molecules from the local database Explorer window onto the main Vector NTI window or into other Vector NTI Advance applications
- Drag molecules or other database objects into subsets
- Open Windows Notepad and drag/drop a molecule on its window. The molecule is displayed in the text format you chose for file system export. See Copy Molecule(s) Tabs (DNA/RNA or Proteins Molecules Table) on page 193.

The following is the list of useful Vector NTI Explorer keyboard shortcuts:

<table>
<thead>
<tr>
<th>Shortcut</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>F2</td>
<td>Rename</td>
</tr>
<tr>
<td>F3</td>
<td>Search</td>
</tr>
<tr>
<td>F4</td>
<td>Select table</td>
</tr>
<tr>
<td>F5</td>
<td>Refresh</td>
</tr>
<tr>
<td>F6 or TAB</td>
<td>Switch to another pane</td>
</tr>
<tr>
<td>ENTER</td>
<td>Open selected object</td>
</tr>
<tr>
<td>CTRL+ENTER</td>
<td>Show Properties</td>
</tr>
</tbody>
</table>

*Table 17.2 Vector NTI Explorer keyboard shortcuts*
Open/Save As Dialog Boxes

Many features of Vector NTI Explorer are available in the Open and Save As dialog boxes (Figure 17.29) appearing in many situations such as when you open a Molecule Display window or save a modified molecule to the database. These dialog boxes display object lists in the same way the Explorer does. You can:

- select between Small Icons, List and Details view modes
- choose object fields to be displayed in the Details mode
- sort by any one of displayed fields
- display the Object Properties dialog box for selected objects.

Select the view modes with the buttons on the upper right or from the shortcut menu. Since there are no menu bars, select Properties (with an object selected), Options or Columns from a shortcut menu from an object or column header (in the Details view). You can also use the Properties button ( ). The Columns page is managed as described under Details View Tab on page 192.

Contact Manager

Vector NTI is increasingly used as a communication tool for exchanging information between researchers. Contact Manager is a tool for creating, updating, deleting and organizing contact records.
To launch Contact Manager, select Database > Contact Manager:

![Contact Manager dialog box](image)

The Contact Manager dialog box displays a list of all stored contact records (Figure 17.30). Many user interface features of the Contact Manager are similar to those of the Open/Save As dialog boxes described above. You can choose record properties to be displayed, sort by any one of the displayed properties and display the record properties dialog for selected objects. For operations on the Contact List, press one of the following buttons:

Add New or Edit Contact Records

- Click the New or Edit buttons to open the Contact Information dialog box (Figure 17.31), where you can enter new data or edit the data for a current contact. Only the Name box is mandatory for the new contact record. All other fields may be left empty and edited later if you don’t know the appropriate data.

Choose OK to save the contact record and return to the Contact Manager.

- Delete deletes the selected contact record from the list. Confirmation is requested before the contact record is deleted.
- Rename changes the name of the selected contact record in an edit box.
Properties ( ) displays the properties of the selected contact record. Only indexed contact fields—Telephone and E-mail—are displayed in the Properties dialog box. Use the Edit operation to view/edit the entire contact record.

User Field Manager

The User Field Manager dialog box maintains a list of user field definitions necessary to provide an appropriate description of the object. Some interface features of the User Field Manager are similar to those of the Contact Manager just described. You can add User Field Definitions, select properties to be displayed, sort by displayed properties and modify User Field Definition properties.

Tip: To walk through a detailed exercise using the User Field Manager to create a new field definition and apply its values to a group of database objects, see Tutorial: Database Explorer in Chapter 7.

Select Database > User Field Manager, opening the User Field Manager dialog box (Figure 17.32):

![User Field Manager dialog box](image)

The operation buttons in the upper left corner perform various operations, described below, on the User Field Definitions. Select the User Field Definition in the list, then press the appropriate button. The add New operation does not require any items to be selected.

Note: You cannot use the exact name of a system field (see the System Fields Table on page 168) as the name of a new User Field you define. Vector NTI field names are case-sensitive. If you really want to name your field ‘Default’, you could use ‘default’ or ‘DEFAULT.’

- Add New ( ) or Edit ( ) (with a field selected) opens the User Field Definition dialog box (Figure 17.33) where you can create or edit User Field Definitions.
Vector NTI Advance 10 User’s Manual

Figure 17.33 User Field Definition dialog box, where you can create or edit User Field Definitions

- **Name**: In the text box, enter (or edit) the name of the new user field. The user field will appear under this name in the descriptions of database objects and in the Database.

- **Type**: Select the data type that the user field will accommodate from the data type choices listed in the drop-down menu:
  - String - arbitrary text string, including numbers and symbols (e.g. a word, sentence or description)
  - Yes/No - binary field which may be set only to “Yes” or “No”
  - List of Strings - list of arbitrary text strings of any length
  - One of Predefined Strings - field which may be set only to one of predefined text strings from a user-defined list
  - List of Predefined Strings - field which may be set only to any subset of predefined text strings from a user-defined list
  - List of URLs - list of web addresses.

- **Cached Fields**: Any User Field can be indexed and “cached” in the database tables. Cached fields are stored in a special memory subsystem from which they can be more quickly recalled. To cache a User Field, check **Cache in the Index Table**. Because cached fields occupy more RAM in your computer, we recommend caching only frequently used fields, for instance, those displayed in the Database Explorer in the Details mode of the Vector NTI Explorer or in the Open/Save As dialog boxes.

If Vector NTI Explorer slows down when displaying an object’s fields in the Details View mode, open User Field Manager, select the field, press the **Edit** button and cache the field. Because cached and ordinary fields have no functional differences, you can turn this option on or off any time you need it.

- **Default Values**: Any User Field may have a default value or values that are displayed when you attempt to fill the previously blank field of the database object. To change the default, click the **Change** button.

- **Predefined Values**: This section is enabled only for fields of the One of Predefined Strings and List of Predefined Strings types. For those fields, enter one or more value strings to choose from. These values are displayed as a list of available options when you try to fill or change the actual object’s fields. Use the text box and the **< Add** and **> Delete** buttons to define the predefined values for the User Field.

- **Description**: Enter a description of the field to help you fill or edit the value of the field in a database object. The description entered here is displayed in the field value dialog box.
for this field as well as the popup label displayed whenever you pause the cursor on the field name in the User Field’s tab of object editors (see below).

**Note:** To save space on your computer’s screen and still allow you to see which predefined values are selected, you may define these values using the ‘abbreviated’ format: abbrev=full form as shown below (these are some of predefined values for a sample “Commercial Sources” field for enzymes):

- A=Amersham Life Sciences-USB
- B=Life Technologies Inc, Gibco-BRL
- C=Minotech Molecular Biology Products
- D=Angewandte Gentechnologie Systeme

Vector NTI displays only abbreviated form (for example A, B, D) in the Explorer windows and other places, but you will see the full forms in value-editing dialogs.

Click **OK** to return to the User Field Manager.

When you have edited data, Vector NTI automatically rebuilds the database indices.

**Back in the User Field Manager:**

- **Rename** lets you change the name of a selected User Field. In the enabled text box, change the field name. **After renaming a User Field, Vector NTI automatically rebuilds the database indices.**

- **Delete** deletes a selected User Field Definition from the Field Definitions list and, optionally, delete field values from database objects. Confirmation is requested before deletion.

- You are also asked to confirm whether to delete the appropriate User Field values from all database objects. Deleting only the User Field Definition does not affect consistency of the database information. When you delete the Field Definition from the list, Vector NTI no longer prompts you to fill this field in new objects and existing objects where it was blank. However, unless you delete the field value from database objects, Vector NTI still stores and displays the existing field data in the database objects. If you try to edit the existing value of that “unlisted” field in any database object, Vector NTI will ask you to add the definition of the field back to the User Field Definition List before proceeding.

- **Properties** display the properties of the selected User Field Description. These properties can be also displayed in the User Field Manager dialog box by opening the shortcut menu from the list headings and selecting Columns.

**Import/Export**

To simplify the maintenance and sharing of User Fields across several databases and/or several computers, you can save your User Field Definitions as “archives.”

To export User Field Definitions to the archives, click the **Export** button. In the dialog box that opens, select the definitions to be exported and click **OK**. In the Export User Field Definitions To dialog box, locate the directory and archive file name for storing the exported definitions. Click **Save** to create the archive.

To import User Field Definitions from the archive, click the **Import** button. In the dialog box, locate the directory and archive file to be read. Click the **Open** button, initiating the import. After importing, Vector NTI automatically rebuilds the database indices.

**Set In/Clear In**

In the User Field Manager, you can assign or “set” a value or clear all values for a particular User Field in a group of database objects.

- **To set a value for a user field:**
Select the user field in Field Definition List (large window). In the database tree in the lower right corner, select a subset of database objects that the field would pertain to. You may choose all database objects regardless of their type or a particular subset for database objects of a particular type (molecules, enzymes, oligos or gel markers). If you don’t want to overwrite existing values of the User Field in the selected group of objects, check the Do not overwrite... box.

In the Global Operations section, click the Set <user field > In button. In the dialog box that opens, specify the field value string to be set and click OK. Vector NTI shows a description of the action about to be performed, asks for your confirmation, then inserts the specified value into the appropriate objects in the subset you selected.

- To clear all values of a user field:
  Select the user field. Choose a subset of database objects in the database tree. You can choose all database objects regardless of their type or a particular subset for database objects of a particular type (molecules, enzymes, etc.)

  Click the Clear <user field > In button. Vector NTI warns you of the action about to be performed. If you confirm the operation, it clears the values of the specified user field in the specified objects.

To set or clear fields in a set of individual objects not currently represented by an existing subset, close the User Field Manager, create the required subset in Database Explorer and return to the User Field Manager. Alternatively, you may change field values in individual objects by editing them in the Explorer one by one as described below.

**Editing User Fields in Individual Objects**

When a User Field is defined in the User Field Manager, you can edit it for any database object using the User Fields tab in the Edit <object name > dialog box (Figure 17.34). To open the Edit dialog box from Database Explorer, select the object in the Object Pane and click the Edit button or select Edit > <Object Name > or the corresponding command from the shortcut menu.

**User Fields Tab in the Edit Dialog Box**

![Figure 17.34 User Fields Tab in the Edit dialog box](image-url)
All of the tabs of this dialog box except the User Fields tab are discussed in the section *Operations on Database Objects* in this chapter. User Fields shows a list of all User Fields currently applied to the object being edited and the list of blank fields defined in the local User Field Definition list but not yet specified for the current object.

To set or edit any field, double-click the field name or select the field name and press the **Change Value** button. If the field you want to edit is not listed in the local User Field Definition list, enter the field definition first. When added, the appropriate field value dialog box is displayed. To quickly clear any User Field of the object, select the name of the field and press the **Remove Value** button.

**User Fields in Molecule Display Windows**

You can view and edit User Fields tab from Molecule Display windows. To open the Edit dialog box, select **Edit > <object > Properties** or in the Text Pane, select Molecule Properties on the shortcut menu opened from the topmost folder <Object Name > or click on the ( ) box to the left of the <Object Name > folder.
A Molecule Display window displays text information, including a description of a molecule, features and a graphical map of the molecule and its sequence. For DNA/RNA molecules you can also see restriction sites, ORFs, motifs, PCR analysis, etc. The Molecule Display window includes powerful graphical tools allowing you to:

- manipulate and analyze DNA/RNA and protein molecules
- visually select fragments of DNA molecules for recombination
- prepare publication-quality figures

A Molecule Display window enables editing in two modes:

- Molecule Editing mode: you can modify a molecule’s functional features and restriction map and sequence. Molecule editing is covered in Chapter 6.
- Picture Editing mode: you can modify the graphical formatting of the molecule’s image, allowing you to create publication-quality illustrations.

Molecules can be saved in the Vector NTI Database or as Molecule Documents. Molecule Document files contain all the information from Molecule Display windows, including the molecule analyses results, as well as graphics display settings you may have formatted or modified for that particular molecule. Molecules in the database maintain consistency in their internal files and graphical representation.
Vector NTI Display Window Options Dialog Box

To view or modify Vector NTI display window options, select Edit > Options in the Vector NTI main workspace or Molecule Viewing window. This opens the Options dialog box (Figure 18.1):

![Options dialog box](image)

Figure 18.1 Vector NTI Options dialog box

The Options dialog box has five different tabs for setting various Vector NTI parameters:

- **General** – sets Vector NTI start-up/exit, date/time format and molecule display profile opening options

  **Note:** The Vector NTI Database Explorer can be configured to open automatically with the Vector NTI main workspace or Molecule Viewing window each time a Vector NTI session is initiated. On the General tab of the Options dialog box, select **Open Local Explorer At Startup**.

- **Undo/Redo** – sets Graphics and Sequence Pane tracking and undo buffer size parameters

- **Molecule Graphics** – sets graphics, font and label formatting options

- **Sequence Editor** – sets block format in the Molecule Sequence Editor dialog box

- **Internet** – sets Vector NTI HTML pages creation parameters

All settings saved in the Vector NTI Options dialog box are preserved between Vector NTI sessions.

Opening Molecules in Molecule Display Windows

The display window can open for a molecule from the Vector NTI database, from a Molecule Document File or from a file in a public format recognized by Vector NTI: GenBank/GenPept, EMBL/SWISS-PROT, BLAST, and FASTA.

Molecule Database Files

Molecules are stored in the Vector NTI database as GenBank/GenPept, EMBL/SWISS-PROT, BLAST, or FASTA files.
Molecule Document Files

Molecule Document files saved in Vector NTI contain all the information contained in a Molecule Display window such as graphical representation and molecule analysis results as well as the molecule itself in a portable format suitable for information exchange. Because the Molecule Document format is based on GenBank/GenProt, any other program able to import GenBank/GenProt files could import Molecule Documents.

All molecule document files, imported and even those created within Vector NTI, are considered external files. Vector NTI restricts the set of operations you can perform on molecules opened from external files. To be able to perform all molecule operations on a molecule from a Molecule Document, you will be prompted to first save the molecule in the Vector NTI database.

When opening a molecule from an external file, you can view it in a Molecule Display window before saving it into the Vector NTI database. You still need to save the molecule in the database to analyze it, construct other molecules from it, etc.

Opening a Molecule from the Database

- From Vector NTI’s Database Explorer:
  Select the molecule table and subset in the upper left corner. In the Database Objects Pane, scroll to the molecule and select `<molecule type>` > Open on the Main Menu bar or double-click on the molecule name in the list.

- From the Vector NTI workspace or a Molecule Display window:
  Click the Open Document button( ) on the main toolbar or choose File > Open. In the Open dialog box (Figure 18.2), select either the Database DNAs/RNAs or Database Proteins tab.

![Figure 18.2 Database DNAs/RNAs tab of the Open dialog box](image)

From the Look In drop-down menu, select the subset where your molecule is located, scroll to the molecule, select it and double-click on it or click OK.

The Properties button ( ) in the Open dialog box is to modify the default set of parameters used to create the Molecule Display window contents. See page 215 for details on Molecule Display Setup.

**Note:** The number of data fields displayed for objects in the Open dialog box can be modified. Select Options from the shortcut menu associated with this dialog box. In the Options dialog box that opens, select fields from the Available Fields list, press the ( ) button to move them to the Dis-
played Fields list. For more information on using this dialog box, refer to page 192 in Chapter 17.

**Opening a Molecule from a Molecule Document File**

To open a molecule from a Molecule Document file, click the Open Document button on the main toolbar or choose File > Open. In the Open dialog box, select the Molecule Files tab. Browse to the location of the molecule file, select it and click OK.

**Opening a Molecule from a Molecule Shortcut File**

Molecule Shortcut files contain the names of database molecules along with the Display Setup parameters. When you open a database molecule via the Shortcut file, Vector NTI uses the Display Setup parameters stored in the file.

To open a database molecule via a Shortcut file, choose File > Open Shortcut. Select the name of the Shortcut file from the file list. To set up a Molecule Shortcut file, see page 232.

**Elements of a Molecule Display Window**

A Molecule Display window is divided into three panes: a Text Pane (upper left), a Graphics Pane (upper right) and a Sequence Pane (bottom) (Figure 18.3).

The Molecule Display window title bar shows the name of the open molecule. Vertical and horizontal split bars separating the panes can be moved using the click and drag technique to resize the viewing areas.

Each pane has scroll bars when the pane contents exceed the viewing area. These allow vertical and horizontal movement through the pane.

The Text Pane has folders containing descriptions and analyses of the molecule currently open. The Graphics Pane displays a graphical depiction of the molecule map. Features, restriction sites, motifs, etc. are shown by symbols. Each symbol typically has a label connected to it by a dropline.
The Sequence Pane of a DNA/RNA molecule depicts the nucleotide sequence with selected functional signals, restriction sites, motifs, ORFs, and protein translations. The Sequence Pane of a protein molecule displays the amino acid sequence and selected features.

Menu Bar and Toolbars

The menu options for the Molecule Display window are discussed when corresponding operations are described. Many menu options have corresponding toolbar buttons and shortcut menu commands.

Many objects and features have associated shortcut menus opened with a right click on the item.

The Main Toolbar remains fairly consistent through Vector NTI Advance, while other toolbars vary according to the display window and the active pane. Because Vector NTI has many toolbars corresponding to different display windows, active panes and applications. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

The Text Pane

The Text Pane of a Molecule Display window contains a thorough description of the molecule, including common database information, Feature map, and results of analyses as well as molecule type-specific information such as restriction maps, ORFs, motifs for DNA/RNA molecules and physiochemical analyses. The content of the Text Pane depends on a specific Display Setup.

Folders

The Text Pane is divided into a hierarchy of folders and subfolders. A folder and its subfolders are called a branch. Text Pane can contain the following folders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Description</td>
<td>Molecule name, length, notes, etc.</td>
</tr>
<tr>
<td>Standard Fields</td>
<td>All standard field descriptions in the molecule’s file: for example, keywords, original source, public database IDs.</td>
</tr>
<tr>
<td>Component Fragments (DNA/RNA)</td>
<td>Describes the fragments from which the molecule is constructed</td>
</tr>
<tr>
<td>Design Description (DNA/RNA)</td>
<td>Design plan determined by Vector NTI’s built-in biological knowledge</td>
</tr>
<tr>
<td>References</td>
<td>Literature references applying to the molecule</td>
</tr>
<tr>
<td>Feature map</td>
<td>Feature map of the molecule</td>
</tr>
<tr>
<td>Restriction Map/Methylation Map (DNA/RNA)</td>
<td>Restriction enzyme cutting and methylation site information for the molecule for those enzymes chosen for display</td>
</tr>
<tr>
<td>Open Reading Frames (DNA/RNA)</td>
<td>ORFs found on the molecule</td>
</tr>
<tr>
<td>Motifs (DNA/RNA)</td>
<td>Motifs found on the molecule</td>
</tr>
<tr>
<td>PCR Analysis (DNA/RNA)</td>
<td>Results of PCR analysis</td>
</tr>
<tr>
<td>Sequencing Primers (DNA/RNA)</td>
<td>Results of sequencing primer search</td>
</tr>
<tr>
<td>Hybridization Probes (DNA/RNA)</td>
<td>Results of hybridization probes analysis</td>
</tr>
</tbody>
</table>

Table 18.1 Text Pane folders on the Molecule Display window
Note: Among the different types of molecules present in the Vector NTI Database are Invitrogen™ (™) Proprietary Molecules. One hundred of the most commonly used, fully annotated, Invitrogen vectors are available in the database. When working with these specific molecules, included in the Text Pane is a hyperlink to the Invitrogen website providing you with relevant information, annotated features and related reagents pertaining to that molecule.

Opening and Closing Text Pane Folders
The general rules for opening and closing Text Pane folders and subfolders are summarized as follows:

- Click on the + at the left of the folder name to open it; click on the – to close it.
- Double click on a folder name to open it; double click again to close it.
- Click on the Expand Branch button ( ) to open it; select View > Expand All to open a folder and all of its subfolders. Click on the Collapse Branch button ( ) to close all related folders.

Finding Molecule Features on Graphics and Sequence Panes
To display molecule features described in the Text Pane in the corresponding region of the molecule in the Graphics and Sequence Panes, open the folder containing the feature and select the feature by clicking on its name. Click the Find button ( ). The corresponding region is marked with a wire-frame in the Graphics Pane and highlighted in the Sequence Pane.

The Graphics Pane
The Graphics Pane graphically displays molecule maps, including features, restriction sites, and motifs, and user annotations. Features represented as clockwise arrows appear on the direct DNA strand; counter-clockwise arrows represent features on the complementary DNA strand.

Circular and Linear Displays
Circular DNA molecules can be displayed in either circular (Figure 18.4) or linear form (Figure 18.5). Linear DNA/RNA molecules and proteins can only be viewed in linear form.
Choose circular/linear display format with the **Circular Display** (>Create Circle<) or **Linear Display** (>Create Linear<) buttons on the toolbar of the Graphics Pane or with the equivalent commands in the **View** menu.

**Figure 18.4** DNA molecule displayed in circular form

**Figure 18.5** DNA molecule displayed in linear form

**Standard Arrangement**

Standard Arrangement sets all symbols to standard sizes, sets labels to a standard font size, and arranges text to prevent or to minimize overlap. When a display window opens, it appears in its default Standard Arrangement settings. After modifying graphics formats, you can return a Graphics Pane to its Standard Arrangement by clicking the **Standard Arrangement** button (Create Graph<) on the Graphics Pane Window toolbar or choosing **View > Standard Arrangement**. While certain changes you have made to the layout in the Picture Editing mode such as label repositioning are lost, many object attributes such as colors, patterns, and font as well as relative symbol sizes are unaffected. Hidden objects remain hidden and the way objects overlap is not affected.
Setting up the Standard Arrangement
To change the sizes of symbols and labels in the Standard Arrangement, press CTRL + the Standard Arrangement button ( ) on the Viewer toolbar, opening the Arrangement Setup Palette:

The palette has three sets of buttons:

- **Stretch Symbols** ( ) and **Shrink Symbols** ( ) buttons respectively increase and decrease the standard thickness of symbols by ~15%. Hold down the SHIFT key while pressing a button to change thickness in small increments.
- **Stretch Labels** ( ) and **Shrink Labels** ( ) buttons respectively increase and decrease the standard size of labels by ~15%. Hold down the SHIFT key while pressing a button to change size in small increments.
- **Spread Labels** ( ) and **Condense Labels** ( ) buttons respectively spread out and compress labels on the linear layout. Hold down the SHIFT key while pressing a button to move labels in small increments.

Click the Standard Arrangement button again to close the palette.

Selecting Signals and Sites
In the Graphics Pane, elements of the functional and restriction maps, ORFs, and motifs can be selected, enlarged or reduced and have various operations performed on them. In brief, to select regions or sites, use the left click + drag method, or click on the feature or its label. Selected regions are marked with a wire-frame.

**Note:** For detailed instructions on selection and viewing options, refer to Chapter 3.

Viewing a Selected Molecule Region
Although a display window always shows an entire molecule, the window can be configured to show only a fragment of an opened molecule. This is useful for manipulating a small region of a larger molecule.

To view only a selected region, activate a Graphics or Sequence Pane and select the targeted region. Click on the View Molecule Fragment button ( ) or select View > View Selection.

Only the selected region of the molecule is now displayed in the Graphics and the Sequence Panes. The list of functional signals, restriction sites and motifs in the Text Pane now reflects only those objects located inside the selected region.

To return to the view of the entire molecule, click on the View Molecule Fragment button again or choose View Whole Molecule from the drop-down menu.

Multi-component Features Display
A multi-component feature is a feature that contains introns and exons. The example shown in Figure 18.6 has two exons and one intron, shown graphically on a molecule. The ordinary feature is shown as a wide arrow on the Graphics Pane. Coding sequences, or exons, are shown
as filled parts of the arrow, while the intervening sequence, or intron, is shown as a gap within the arrow.

![Fragment of CEF26D2](image)

Figure 18.6 Multi-component feature with two exons and one intron

Selecting and Translating Exons

When a DNA molecule containing a non-contiguous multi-component CDS is open in a molecule display window, the user can translate the exons three ways. Click on any of the exons to select the entire CDS. **Translate commands are enabled on the shortcut menu.**

The first two options create a new protein. The last simply displays the translation of the selected area in the Sequence Pane.

- **Translate Feature into New Protein**
  This command translates the entire feature into a new protein molecule. Because reading frames may be shifted in translating intron/exon features in this method, the resulting amino acid sequence may not match the valid protein sequence for intron/exon features translated in this manner.

In the above case, a new molecule display window opens featuring the new protein.

- **Translate Feature in Sequence Pane**
  This command displays the translation of the selected feature in the Sequence Pane of the Molecule Viewing window. A new protein molecule is NOT created.

For more information on Translation, see Translation, page 292. For information on creating a new molecules by splicing together the exons of a multi-component feature, see Creating New DNA/RNA or Protein Molecules Using Splicing, page 307.

**Changing the Settings for Graphics Pane**

For changing Graphics Pane settings, see Display Setup later in this chapter. For editing a molecule or modifying its graphical format in the Graphics Pane itself, refer to Molecule Editing in Chapter 20.

The Sequence Pane

A Sequence Pane displays a molecule’s nucleotide or amino acid sequence as well as selected features.

<table>
<thead>
<tr>
<th>Features</th>
<th>Display</th>
<th>Position on Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional features (labeled)</td>
<td>deep red serrated lines</td>
<td>Above—direct strand; protein Below—complementary</td>
</tr>
<tr>
<td>RENs (labeled)</td>
<td>blue wavy lines</td>
<td>Above only</td>
</tr>
<tr>
<td>Motifs (labeled)</td>
<td>deep red serrated lines</td>
<td>Above—direct Below—complementary</td>
</tr>
</tbody>
</table>

Table 18.2 Selected features of Sequence Pane
Selective display of these features is defined by corresponding commands under the Analyses menu option. They can also be defined in Display Setup, discussed later in this chapter.

**Changing the Settings for Sequence Display**

To change the settings for sequence display, select parameters in the Display Setup dialog box, discussed in detail on page 215.

**Setting an Internal Start Point in the Sequence Pane**

By default, Vector NTI numbers the molecule sequence in the Sequence Pane starting with the number 1, which coincides with the origin of the molecule.

You can create an additional, secondary numbering system in the Sequence Pane display (DNA/RNA Molecules only) by setting an internal start point. The sequence is numbered starting with 1 at the internal start point, with increasing numbers to the right of the start point and decreasing numbers (negative numbers) to left of the start point.

To set an internal start point, choose View > Internal Start Point from the drop-down menu in a DNA/RNA Molecule Viewing window. The Internal Numbering dialog box displays (Figure 18.7):

![Internal Numbering dialog box](image)

Enter the number of the base where you want to set the internal start point or press the Current button to set the number to the current cursor position and then press the OK button. Internal Start Point numbering is displayed in the Sequence Pane:

![Internal Start Point numbering in the Sequence Pane. Internal start point is circled.](image)

**Note:** Internal start point numbering is considered a form of molecule editing and can, therefore, be saved as part of the molecule.

**Navigating the Sequence**

In the Sequence Pane, the cursor becomes an I-beam that you can move with the mouse or the arrow keys. Pause the cursor over a sequence position for a few seconds, and a pop-up label identifies its exact position on the sequence.
Selecting Regions of the Sequence

There are several methods for selecting a sequence in the Sequence Pane:

- Use the LEFT-CLICK + DRAG technique, dragging the cursor across the nucleotides.
- Select Edit > Set Selection. Enter the residue range in the dialog box that opens.
- Click on the Set Selection box on the Status Bar, opening the Set Selection dialog box. Enter the residue range in the dialog box.

The selection range is always displayed in the Set Selection box on the Status Bar.

The general rule for modifying sequence selections is to hold the SHIFT key (to alter the selection by small increments) or the SHIFT + CTRL keys (to alter the selection by large increments) as you press the left and right arrows.

A selected fragment in the Sequence Pane is marked simultaneously with a wire-frame in the Graphics Pane.

A fragment selected in the Sequence Pane is marked simultaneously with a wire-frame in the Graphics Pane.

Translations

Translating in the Sequence Pane

Select the region you wish to translate.

To translate in the selected frame in the Sequence Pane, press the Translate Direct Strand or Translate Complementary Strand buttons in the Window toolbar, or choose the corresponding commands from the Shortcut menu.

If ORFs are displayed, you can translate ORFs in the Sequence Pane by choosing Translate ORFs from the Shortcut menu.

Notes:

- Translations of the direct strand appear above the sequence; translations of the complementary strand appear below the sequence.
- You can select three- or one-letter amino acid codes through the Display Setup dialog box, discussed on page 215 in this chapter.

To clear Sequence Pane translations, do one of the following:

- click the Erase Translations button
- choose View > Clear All Translations from the drop-down menu

A variety of Sequence Pane translation options, including those mentioned above are available under the Analyses drop down menu by choosing Analyses > Translation > In Sequence Pane > <translation option>. For more information on these options, see Translating in the Sequence Pane, page 293.

Translating into a New Protein Molecule

You can also translate a DNA/RNA molecule Sequence Pane selection into a new Protein molecule stored in the Vector NTI Database. The selected portion of a DNA/RNA molecule can be translated (direct or complementary strand) and saved in the Vector NTI Database by choosing the File > Create New Sequence > Using Translation Procedure (Protein) > <strand option> from the drop-down menu in the Vector NTI Molecule Viewing window.

Other options for translating DNA/RNA selections into new protein molecules are available under the Analyses drop-down menu by choosing Analyses > Translation > Into New Protein.
> <translation option> For more information on these options, see Translating into a New Protein Molecule, page 293.

Setting the Genetic Code

In Vector NTI, Standard genetic code is used for translations by default. However, you can choose from among a list of alternate genetic codes or create your own custom genetic code to use for translations. For instructions on setting the genetic code, see Setting the Genetic Code for DNA/RNA Translations, page 294.

Phase Indicators for ORFs and Protein Translations

Phase indicators in the left margin of the Sequence Pane show the phases of ORFs and protein translations. For translations of the direct strand, the phase is determined by the position of the first selected nucleotide relative to the beginning of the molecule. For the complementary strand, the phase is determined by the position of the last selected nucleotide relative to the end of the molecule.

Finding DNA/RNA Sequence Fragments

To quickly find a defined fragment on a sequence, in the Sequence (or Graphics) Pane, click the Find button ( ), select Edit > Find Sequence or press CTRL + F. In the Find Sequence dialog box (Figure 18.9), enter the desired sequence, indicate the strand to search, choose exact or degenerate matching and set the acceptable mismatch tolerance. The sequence can contain any standard IUPAC symbols for nucleotide sequences. (See Appendix C.)

Figure 18.9 Find Sequence dialog box (DNA)

Press the Find Next button. The search begins from the current caret position or the current selection. A located fragment is automatically selected in the Sequence Pane. If none are found, the caret does not move from its current position.

Finding Protein Sequence Fragments

To quickly find a defined fragment on a sequence, in the Sequence (or Graphics) Pane, click the Find button ( ), select Edit > Find Sequence or press CTRL + F. In the Find Sequence dialog box (Figure 18.10), enter the desired sequence and set the acceptable mismatch tolerance.
The sequence can contain any standard IUPAC symbols for amino acid sequences. (See Appendix C.)

Click the **Find Next** button. The search begins from the current caret position or the current selection. A located fragment is automatically selected in the Sequence Pane. If none are found, the caret does not move from its current position.

**Change Text Character Display Attributes**

Character display attributes in the Sequence Pane can be changed in a manner as you would in a word processor program. To change the font name, font size, font color and font attributes (bold, italic, underline) for any character in a nucleotide sequence, highlight these characters and click the corresponding buttons on the toolbar. For further information, refer to Chapter 3.

To preserve text formatting between sessions, save the Vector NTI molecule as a Molecule Document.

**Cut, Copy and Paste**

To Cut, Copy, or Delete a region of the molecule’s sequence, select the region and choose the appropriate command from the Edit menu. *If a region to be deleted overlaps with features, a confirmation dialog box opens, allowing you to keep or remove affected features.* All the features inside the deleted region will be removed without confirmation; positions of other features will be adjusted accordingly.

A copied sequence is placed on the Clipboard in the standard text format, so you can freely exchange the sequence data between Vector NTI and other applications. All sequence formats recognized by Vector NTI can be pasted from the Clipboard. Sequence formats are described in detail in Chapter 14 (Import section).

To paste a sequence from the clipboard, set the cursor to the insert position. Choose **Edit > Paste Sequence**, opening and displaying the sequence in the Insert Sequence dialog box. Click the **OK** button, and the sequence is inserted at the cursor position. If the insertion point is within a feature, a confirmation dialog box opens, to let you keep or remove the affected features. The positions of all features surrounding the insertion point are updated automatically.

Insert a new sequence fragment manually by setting the cursor at the insertion point; choose **Edit > New > Insert Sequence**. In the Insert Sequence dialog box, enter the sequence to be inserted. For more information on editing a sequence, see Chapter 19.

**Editing a Molecule Sequence**

Instructions for editing a molecule sequence and its text attributes are found under Molecule Editing in Chapter 19.

**Molecule Display Setup**

Vector NTI uses Display Setup parameters to create the Molecule Display window contents. You can edit Display Setup before or after a display window is created, assign names to different settings and store them as a Setup profile, and change the default Setup settings.
To open the Display Setup dialog box, click the **Display Setup** button ( ) on the Window toolbar. Select **Display Setup** from the drop-down menu.

The Display Setup dialog box has two forms, one for DNA/RNA molecules and another for proteins. The type of Molecule Display Setup dialog box that opens corresponds to the type of molecule display window from which it was launched (i.e. If you launch the Display Setup dialog box from a DNA/RNA Molecule Viewing window, the DNA/RNA Molecule Display Setup dialog box opens; if you launch the Display Setup dialog box from a Protein Molecule Viewing window, the Protein Molecule Display Setup dialog box opens.

**DNA/RNA Display Setup**

The following figure (Figure 18.11) is the Display Setup dialog box for DNA/RNA Molecules:

![Display Setup dialog box for DNA/RNA molecules](image)

Figure 18.11 Display Setup dialog box for DNA/RNA molecules

To apply a previously saved Setup Profile to a current display window, select the Profile in the Setup Profile drop-down menu. Click **OK**.

Use the check boxes to indicate the information to be displayed in the Molecule Display window. Click on the **Setup** buttons (described below) to open corresponding dialog boxes where you can specify in more detail what information should be displayed for each category. Click **OK** in each to return to Display Setup.

**Feature Map Setup**

Check the **Feature Map** check box and click the **FMap Setup** button to select the types of features to be displayed in the Display window’s Sequence Pane. Some features appear by default in the Text and Graphics Panes. The Feature Map Setup dialog box appears (Figure 18.12):

![Feature Map Setup dialog box](image)

Figure 18.12 Feature Map Setup dialog box
Click the `<Add>` button to select one or more of available feature types. Click the > Remove button to remove previously selected feature types or >> Remove All empties the entire list.

You can limit the features shown in the Sequence Pane according to feature length. To set a feature length limit, enter the maximum feature length in bases in the textbox in the lower left of the FMap Setup dialog box. Leave the textbox blank for no limit.

**Motifs Setup**

Check the Motifs checkbox and press the Motifs Setup button to open the Motifs Setup dialog box (Figure 18.13) that controls the motifs map display.

![Motifs Setup dialog box](image)

The list box in the upper part of the dialog box contains motifs already chosen for a search.

- **Add New** – Adds new motifs and Edit lets you edit a highlighted motif. Both of these operations take place in the Oligo Editor dialog box.
- **Oligo Database** – Uses the Vector NTI oligo database as a source for motifs to be searched for.
- **Oligo List** – Loads oligos from the Oligo List.
- **Remove** – Removes selected motifs.
- **Remove All** – Removes all listed motifs.
- **Save To Database** – Saves the selected motif to the oligo database.
- **Edit** – Allows you to edit the selected motif.
- **Analyze** – Initiates an analysis of a highlighted motif.
- **Find Sites** – Press the Similarity button to specify the similarity between motif and molecule sequence for the motif’s site to be accepted. In the Find Motif Locations dialog box (Figure 18.14) that opens, choose the method for calculating the similarity and the similarity requirements.

![Find Motif Locations dialog box](image)
- **With Best Similarity Only** selects only those motif sites that received the best relative score.
- **With Similarity >= Similarity Threshold** displays all motif sites that have a similarity more than or equal to the similarity you specified in the Similarity Threshold box.
- To filter out all the motif sites that do not meet the specific requirement for 3’ end similarity, check the **Last** box and enter the number of nucleotides on 3’ end that should have 100% similarity.

Any motifs that have more sites than specified in the Maximum Number Of Sites box in the Motifs Setup dialog box (Figure 18.13) are displayed as grayed-out folders in the Text Pane; they are not displayed at all in graphics or Sequence Panes.

**Similarity Between Ambiguous Nucleotides** specifies acceptable similarity between ambiguous nucleotides (if any). The **Average**, **Minimum**, and **Maximum** buttons indicate that the average, minimum, and maximum possible similarity will be calculated respectively for any nucleotide pair. For instance, if you are calculating similarity between N and A, then the average similarity is 25%, the minimum similarity is 0%, and the maximum similarity is 100%. In case of R and A they are 50%, 0%, and 100%; in case of R and T—0%, 0%, and 0%. For clarification of these IUB nucleotide symbols, see Appendix C.

<table>
<thead>
<tr>
<th></th>
<th>N-N</th>
<th>N-R</th>
<th>N-A</th>
<th>R-W</th>
<th>R-A</th>
<th>R-T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Average</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Table 18.3 Average, minimum, and maximum possible similarity*

Click **OK** to return to the Motifs Setup dialog box.

- **Sorted By** Select the motifs sorting procedure to be used for the Motifs folder. Choose the **Motif String** radio button to sort motifs alphabetically or the **Number of (Recognition) Sites** radio button to sort motifs by their number of sites on the molecule.
- **Strands to Search** Turn these boxes on or off to specify the strand(s) to be searched.
- **Ignore Outside / Ignore Inside** Enter these regions to specify additional limitations on the motif sites to be displayed. Specify the **Ignore Motifs With Sites Outside Region** to display only those motifs located inside the region. Specify the **Ignore Motifs With Sites Inside Region** to display only those motifs located outside the region.
**Sequence Setup**

Check the **Sequence** checkbox and press the **Sequence Setup** button to indicate how the sequence should be displayed in the Display window’s Sequence Pane. The Sequence Setup dialog box appears (Figure 18.15):

![Sequence Setup dialog box](image)

Set the number of nucleotides per block in the sequence display (10 is default), the number of blocks per line (5 is default) and whether protein translations should be shown in one- or three-letter code (three is default). Check the Prefer Single-Stranded Display box if you prefer the molecule sequence displayed as single-stranded in the Sequence pane. Set the default font name, size, color and effect characteristics.

**Restriction Map Setup**

Check the **Restriction Map** checkbox and press the **RMap Setup** button to specify the restriction sites to be displayed. The Restriction Map Setup dialog box (Figure 18.16) displays in the Use Enzymes column the restriction enzymes, also called RENs, whose sites are shown by default.

To remove enzymes from the display list, select the enzyme and click **Remove**. The **Remove All** button clears the list. Click the **Add** button to select database enzymes to be added.

After clicking the **Add** button, to quickly move to the vicinity of the enzyme of interest in the Choose Database Enzymes dialog box enzyme list, highlight the first enzyme name on the list and, using the keyboard, type the first letter (or first few letters, quickly) of the name of the enzyme of interest. The enzyme list scrolls to the vicinity of that enzyme (Note: For this technique to work, the enzyme list must be sorted alphabetically by enzyme name. To sort the list
alphabetically by enzyme name, click on the Name column header in the Choose Database Enzymes dialog box.

![Restriction Map Setup dialog box](image)

Figure 18.16 Restriction Map Setup dialog box

- The **Sort Site...** radio buttons determine how the restriction sites are displayed in the Text Pane. **Enzyme Name** sorts the RENs alphabetically. By **Number of Recognition Sites** sorts the RENs by the number of their sites on the molecule.
- **Permitted Terminus Types** check boxes filter the types of RENs displayed.
- **Ignore RENs Having Less Than/More Than ... Sites** removes from the restriction map RENs that do not fall within the specified cut site range. Such RENs will be listed but grayed out in the Restriction Map folder in the Text Pane. They will not be displayed at all in the graphics and Sequence Panes.
- **Ignore RENs Cutting Outside Region** causes only those enzymes with sites inside the region to be displayed.
- **Ignore RENs Cutting Inside Region** causes only those enzymes with sites outside the region to be displayed.

**Picture Type Checkbox**

If you check the **Prefer Linear** checkbox in Display Setup, Vector NTI uses linear display for both circular and linear molecules.

**ORF Setup**

Check the **ORFs** checkbox and press the **ORF Setup** button to open the ORFs Setup dialog box (Figure 18.17) for specifying how open reading frames are displayed.

![ORFs Setup dialog box](image)

Figure 18.17 ORFs Setup dialog box
In the **Minimum ORF Size** field, enter the minimum ORF size in codons.

Check the **Nested ORF** check box to look for nested ORFs (ORFs that have the same stop codon but different start codons.)

In **Start** and **Stop Codons** fields, enter start and stop codons for the new viewer’s ORFs. Press the **Default Start & Stop** button to set the start and stop codons to the following conventional values: Start codons—ATG, GTG; Stop codons—TAA, TGA, TAG.

Check the **Include Stop Codon in ORF** box if you want the stop codon to be considered part of the ORF. Otherwise, the stop codon is not considered part of the ORF and is not included.

**Incomplete ORFs**

- **Undefined Start/Undefined Stop**: Check the appropriate box, depending on whether you want to search for ORFs with undefined starts, stops or both.
- **Minimum Size**: Specify the minimum size in base pairs for the incomplete ORF to be displayed.

**Note**: Incomplete ORFs are displayed as a dashed arrow in both the Graphics and the Sequence Pane. Complete ORFs are displayed as solid arrows.

**ORF Names**

- **Old Style**: ORF# (sequential)
- **New Style**: Nucleotide position of ORF start, direct or complementary, sequential #

**Analyses Setup**

Check the **Analyses** checkbox and press the **Analyses Setup** button to open the Analyses Setup dialog box (Figure 18.18). In the Analyses Setup dialog box you can specify the analyses to be shown in the Sequence and Graphics Panes and in the External Analyses folder in the Text Pane.

From the Available Subsets drop-down box, choose the Vector NTI Database subset of saved analyses you want to display. Click the **>>** button to move the selected subset into the Load Analyses Results from Subset section. All analyses listed in this subset are shown in the Molecule Viewing window.
Graphics Display Settings

You can edit the graphical styles for graphical map elements directly from the Display Setup dialog box or in the Graphics Pane in Picture Editing mode. Click the Edit button in the Graphics Display Settings group, opening the Graphics Display Setup dialog box (Figure 18.19).

![Figure 18.19 Graphics Display Setup dialog box](image)

Select the graphical map element type and define the shape and style for a particular symbol picture as well as the style for a label attached to the symbol picture. Add new styles by clicking the More buttons (one button adds label styles, another adds symbol styles). When the styles are defined, click OK.

**Note:** You can also edit existing styles directly in the Graphics Pane of a Molecule Display window as described later in this chapter.

Save To File/Load From File

Display settings configured in the Molecule Display dialog box can be saved to and loaded from external files on your hard drive. To save the current Display Setup profile to an external file, click the Save To File button and enter a name and storage location for the new Molecule Setup file. To load a previously saved Display Setup file, click the Load From File button and navigate to the location of the file on your hard drive. Select the Setup Profile from the dialog box; the selected settings are applied to the current Molecule Display window.

Setup Profile

The upper part of the Display Setup dialog box contains a list of saved setup settings called Setup Profile. Add your newly defined settings to the Profile list by clicking the Save Settings As button. Enter a name for these settings and click OK. In the future, you can apply at once all of the settings saved in a Profile to a new molecule opened in a Molecule Display Window. Open a saved setting by selecting the Profile name in this list or View > Display Profile.

Vector NTI always uses the default Display Setup when you open a molecule from the database in a new Molecule Display window, with two exceptions:

1. As you open a molecule from the Vector NTI workspace window, you can specify the Display Setup for the new display window. To do this, choose File > Open from the Vector NTI window, select the Database tab for the type of molecule you want to open (DNA/RNA or Protein) and click the Display Setup button. In the Molecule Display Setup dialog box that opens, select the Settings Profile you want from the list of saved profiles.

2. If you open a molecule from the database via a Molecule Shortcut file, the settings stored in the shortcut file are used instead of the (Default) entry in the Setup Profile.
Protein Molecule Display Setup

The options in the Display Setup dialog box (Figure 18.20) for a protein differ slightly from that for nucleic acid molecules.

**Feature Map Setup**

With a protein in the Molecule Display window, press the **FMap Setup** button to select the types of features to be displayed in the Display window’s Sequence Pane. (*Many features appear by default in the text and Graphics Panes.*) The Feature Map Setup dialog box for proteins differs from that for DNA/RNA only in the features listed. The functionality of the dialog box is identical to the DNA/RNA Feature Map Setup box. See page 216.

Click the `< Add` button to select one or more of available feature types. Click the `> Remove` button to remove previously selected feature types. Click the `>> Remove All` button to delete the list.

**Motifs Setup**

Click the **Motifs Setup** button to open the Protein Motifs Setup dialog box (Figure 18.21) that controls the protein motifs map display:

**Note:** The list box in the upper part of the dialog box contains motifs already chosen for a search.

Features of the Protein Motifs Setup dialog box (Figure 18.21) their associated subdialog boxes are described in the following bullets:
Add – Launches the All Protein Motifs dialog box (Figure 18.22).

In the All Protein Motifs dialog box, highlight the motifs for which you wish to search and click the OK button to enter the selected motifs into the Protein Motifs Setup dialog box (Figure 18.21).

In the All Protein Motifs dialog box, you can also create new motifs and add them to the motifs list by clicking the Add New button. You can edit an existing motif by highlighting the motif and clicking the Edit button. Clicking either the Add New or Edit buttons in the all Protein Motifs dialog box launches the Protein Motif Properties dialog box (Figure 18.23):

In the Protein Motif Properties dialog box, you can create and edit motif patterns.

To define a new pattern, type the pattern in the New Pattern String box. Click the Add to Patterns button to add the new pattern to the current list of patterns. To remove a pattern from the list, click the Remove button. To remove all patterns from the list, click the Remove All button.

- Remove – Removes selected motifs.
- Remove All – Removes all listed motifs.
- Find Sites – Specify the similarity between motif and molecule sequence for the motif’s site to be accepted. Select the method for calculating the similarity and the similarity requirements.
  - Find Best Fit – Finds only those motif sites that have the best relative similarity.
  - With Similarities Not Less Than – Finds motif sites with similarity equal to or greater than the specified percent similarity. Number of sites found is limited by the Maximum Number of Sites value specified.
- Similarity for Ambiguous AA – Specifies the acceptable similarity between ambiguous amino acids, based on the Minimum, Average or Maximum similarity for any amino acid
pair. For an example of similarity calculations, see the Similarity Between Ambiguous Nucleotides bullet of the DNA Motifs Setup section on page 217.

- **Sorted By** – Select the motifs sorting procedure to be used for the Motifs folder in the Text Pane. Choose the **Motif Name** radio button to sort motifs alphabetically or the **Number of Sites** radio button to sort motifs by their number of sites on the molecule.

- **Ignore Outside / Ignore Inside** Enter these regions to specify additional limitations on the motif sites to be displayed. Specify the **Ignore Motifs With Sites Outside Region** to display only those motifs located inside the region. Specify the **Ignore Motifs With Sites Inside Region** to display only those motifs located outside the region.

**Sequence Setup**

Click the **Sequence Setup** button to open the Sequence Setup dialog box (Figure 18.24):

![Sequence Setup dialog box](image)

In this dialog box, set the number of amino acids per block in the sequence display and the number of blocks per line. Set the default font name, size, color and effect characteristics. A sample of the designated configuration appears in the lower text box.

**Picture Type**

Ordinarily, if you select a protein fragment and click the **View Selected Fragment** button for closer inspection, the selected region “expands” to fill the entire Graphics Pane. If the **Prefer Constant Horizontal Resolution** check box is checked in the Protein Display Setup dialog box, the current protein display is “not scalable,” meaning that the current visual length of graphics features is retained instead of stretching the picture to fit in the window.

**Analyses Setup**

The Display Setup Analyses settings for proteins are managed as explained on page 221.

**Graphics Display Settings**

The Display Setup Graphical Map settings for proteins are managed as explained on page 222.

**Save To File/Load From File**

Display settings configured in the Molecule Display dialog box can be saved to and loaded from external files on your hard drive. To save the current Display Setup profile to an external file, click the **Save To File** button and enter a name and storage location for the new Molecule Setup file. To load a previously saved Display Setup file, click the **Load From File** button and navigate to the location of the file on your hard drive. Select the Setup Profile from the dialog box; the selected settings are applied to the current Molecule Display window.
**Setup Profile**

See setup Profile in the DNA/RNA Display Setup section on page 226.

**Picture Edit Mode: Formatting and Annotating Pictures**

Vector NTI has two modes of operation in Graphics Panes.

- **Molecule Editing mode** (default) in which the Graphics Pane serves as a visual environment for viewing and editing the content of a molecule itself. In molecule editing, you are changing the nature of the molecule map and sequence. Selecting DNA molecule fragments for Construction and Design also take place in Molecule Editing mode. See Chapter 21 and Chapter 22 for details about Construction and Design. Molecule Editing mode is discussed in Chapter 19, Editing Molecule Content.

- **Picture Editing mode** where you can edit the graphical representation of an image such as format and arrange individual graphics symbols and labels. This mode is useful for preparing publication-quality figures of the molecule.

When in Picture Editing mode, Vector NTI changes the way it selects objects in the Graphics Pane: the elementary units of selection are graphics objects like labels and symbols, not elements of the molecule content. In Picture Editing mode, you can reposition and reformat labels and symbols, hide objects to prevent them from being displayed, and format multiple objects at once using centralized style sheets.

**Notes about Picture Editing Mode:**

- If you intend to edit the molecule itself, it is recommended that you perform that first before you format molecule graphics.

- Picture editing changes are not saved in database, but in a Molecule Document file.

To activate Picture Editing Mode, in an active Graphics Pane, click the **Edit Picture** button on the toolbar. When the Graphics Pane is in the Picture Editing mode, the button is shown in depressed state and View > Edit Picture is checked.

All of the following operations are performed in Picture Editing mode.

**Formatting Objects on the Graphical Map**

To format an object in the Graphics Pane, select the object by clicking on it or its label. Review selection techniques in Chapter 3. To format more than one graphics object at a time, HOLD SHIFT + CLICK on each.

Selected objects in Picture Edit mode are displayed with selection boxes or handles at each corner. If this is not the case, check to make sure the **Edit Picture** button is depressed.

When the graphics objects to be formatted are selected, choose **Edit > Properties**, press ALT + ENTER or select **Properties** from the associated shortcut menu.

In the Properties dialog box, select **Line**, **Fill**, or **Text** attributes (if a label is selected). If a single symbol is selected, you may also change its **Shape**.

On the Properties tabs, select the pattern and color for Lines and Fill for selected graphics objects. On the Shape tab, click the down arrow for selections. Click **OK**.

**Note the following about formatting objects:**

- **Lines** associated with symbols are the lines that surround the main symbol. Lines associated with labels are the droplines connecting the labels to its corresponding symbol.

- **Fill** refers to the pattern and color with which the selected symbols are filled. Changing a symbol’s filling does not affect the formatting of the line around that symbol.
The Default Shape option forces Vector NTI to assign the default shape to the given symbol. You can edit the association between the element type and the default shape/style in the Graphics Display Setup dialog box accessible from the Display Setup dialog box. Note: You cannot undo the Shape formatting.

If you have selected label(s), the Properties dialog box includes a Text field (Figure 18.25).

In the text box, enter ordinary text or choose from a variety of special codes that are converted into information about the symbol to which the label is attached. To insert a substitution code, choose the code from the Substitutions box and click the Insert button. The code is inserted at the current caret position.

This substitution model allows assignment of a single text "style" to all labels of a certain kind. For example, if the text attribute for all restriction sites is "@N (@S)" then the actual label for a SmaI site 5354 bp will look like “SmaI (5354)” with the REN name and site position substituted for @N and @S. You might want to experiment with the different substitution models.

Using the Font button on this tab, choose the font, font style, font size, effects, and display color attributes of labels you selected. Attributes of currently selected font are displayed to the right of the Font button.

Note: The font size you choose is not necessarily equal to the font size on the screen or on the printed output. Displayed objects are stretched and shrunk as the picture magnification is changed.

Annotations

To insert a text annotation on a graphical map, click the Add Annotation button in the Window toolbar or choose Edit > New > Add Annotation. In the Annotation dialog box, enter the text of a new annotation label. Click OK to return to the Molecule Display window where the annotation is highlighted and selected with handles. Click and drag it to the location you choose in the Graphics Pane.

You can move annotations and change their font and formatting using the label formatting techniques described above.

Note: Annotations are elements of the graphical representation of the molecule only and are not stored in the database. To save an annotation, you need to save the display window as a Molecule Document. To associate text information with a molecule itself, not its temporary graphical representation, use the Comments or User Fields described in Chapter 17.

Deleting Annotations

To delete an annotation, select it in the Graphics Pane, and choose Edit > Delete Annotation, or to delete an annotation from a Text Pane, click the annotation subfolder to select its line, and
click the **Delete Annotation** button ( ) in the Window toolbar or choose **Edit > Delete Annotation**.

To delete annotations from the Graphics Pane, select them in the Picture Editing mode and choosing **Edit > Delete Annotation**.

**Note:** Annotation subfolders are the only folders of the Text Pane that can be deleted.

### Undo and Redo

You can undo/redo movement and formatting of graphics objects, redefinition of styles, etc. Undo/Redo also works in the Sequence Pane. The list of actions, tracked by undo/redo as well as the size of the undo buffer can be changed in the Undo/Redo tab, accessed from **Edit > Options**.

- To undo an action, click the **Undo** button ( ) in the main toolbar or choose **Edit > Undo <Formatting>**.
- To redo an action, click the **Redo** button ( ) in the main toolbar or choose **Edit > Redo <Formatting>**.

### Changing the Sizes And Shapes of Symbols

To change the size or shape of a symbol, click the **Edit Picture** button ( ); you cannot modify a graphic unless the Graphics Pane is in **Edit Picture mode**. Next, select the object to be changed with a mouse click. Black drag boxes will appear at its corners. Move the cursor to any drag box, changing it to a thin four-headed arrow ( ). Click and drag the black box, causing the symbol to change shape and/or size (Figure 18.26). Release the mouse when the symbol has the new shape you want.

![Figure 18.26 Changing a symbol's shape and/or size](image)

### Moving Objects on the Graphical Map

To move a graphics object around the Graphics Pane, select the symbol or label. When the cursor placed in the middle of the selected object turns into a four-headed arrow ( ), click and drag the object to the new location.

As labels are dragged, their droplines follow, tracking the connection to the object to which they refer.

Most symbols representing features, restriction sites, etc., do not move completely freely around the graphical map. They do move freely perpendicular to the “sequence” bar/ring, but Vector NTI prevents them from moving parallel to the sequence bar without your permission. This is to keep the objects in the places corresponding to their actual locations in the nucleotide or amino acid sequence.

To move a symbol parallel to the sequence bar of a graphical map, press and hold **CTRL+ SHIFT** and drag the object.
**Note:** This does not change the coordinates of the feature in the Feature map. To change the coordinates you need to edit the feature in the Molecule Editing mode described in Chapter 19.

**Changing the Display Order Of Overlapping Graphics Objects**

To change the display order of overlapping graphics objects, select them; choose **View > Picture Element > Bring To Front or Send To Back** or the corresponding commands from the shortcut menu. **To make these changes, the Graphics Pane must be in Edit Picture mode.**

The following image demonstrates the results of bringing the DNA ring to the front of a circular molecule display (Figure 18.27):

![Figure 18.27 Bringing the DNA ring to the front of a circular molecule display](image)

**Hiding Graphics Objects**

To hide an object, select it and choose **Edit > Hide**. *If you hide symbols, their labels remain visible unless you hide them specifically.* The Hide feature is useful if you don’t want objects to appear on a final picture or if you want to have more room for better layout of important picture elements.

**Restoring Hidden Graphics Objects**

To restore hidden graphics objects to the Graphics Pane’s display, click the **Undo** button or **Edit > Undo**. You can also activate the Text Pane, link the text and Graphics Panes, and open the folder(s) containing the hidden graphics objects.

**Styles**

A *style* is a named set of formatting attributes for graphics objects. Styles in Vector NTI work much the same as they do in most word processing programs. Styles provide a convenient and powerful way to organize the formatting of graphical maps.

Each style contains one or more of the following attributes: line, fill, text, and shape.

Although Vector NTI has standard styles for symbols and labels, you can also define your own. Once created, styles can be saved together with other parameters in the Display Setup Profile or in Vector NTI Setup files.

To learn what style a graphics object has, turn on the Picture Editing mode, select the object and select **View > Picture Element > Style** or **Style** from the shortcut menu. The style box in the Picture Element Style dialog box shows the style of the selected object. If more than one object is selected, the style box shows the style of the last object selected. The last selected object is displayed with small squares around it.
Assigning Styles to Graphics Objects

To assign an existing style to a graphics object, select the object(s) to which you want to assign a style and choose View > Picture Element > Style or Style from the shortcut menu. In the Style box in the Picture Element Style dialog box, choose an existing style from the pull-down list, or type a name, and click the OK button.

If you select an existing style or type its name, that style is assigned to the selected graphics object(s), and the object(s) assume its format.

If you type a name that is not in the style list, a new style with the format of the currently selected object(s) is created. If the selection includes more than one graphics object, a newly created style contains only those attributes that are the same for all selected objects.

A style is the base format for a graphics object. You may make additional formatting changes to graphics objects using the methods described in preceding sections.

Redefining a Style’s Format

To redefine the style of a graphic object, select the object, open the Properties dialog box, opened from the shortcut menu and make the desired changes for the object. Click OK. Select View > Picture Element > Style or select Style from the shortcut menu. In the Style drop-down menu, the name of the current style is displayed. Click OK, signaling that you wish to apply the new format to an existing style. In the Reapply Style dialog box (Figure 18.28) that opens, choose the redefine operation you wish to occur and click OK.

![Figure 18.28 Reapply Style dialog box](image)

To accomplish the same end, you could select View > Picture Element > Redefine <name> Style by Example or the corresponding command from the shortcut menu.

If you have chosen the Redefine Style option, any objects (even unselected ones) that currently have that style will automatically change format to reflect the style’s new settings.

If you have chosen the Return Formatting to Style option, then the selected objects’ format is restored to that Style’s format. Edit > Undo or the Undo button restores the former style.

Copy and Paste Styles

Once you have modified a symbol or label with a new format, you can easily reformat other objects with the same format. Select the modified object, and select Copy Style from the shortcut menu associated with the modified object. Move the cursor to another feature or label, and select Paste Style from the shortcut menu.

Saving Styles

The set of the styles and Standard Arrangement parameters appearing in the Molecule Display Window are a part of the Display Setup settings.

To save a display window’s current styles, choose the Edit > Display Setup or click the Display Setup button(n) in the Window toolbar and choose Display Setup from the drop-down menu. In the Display Setup dialog box, there are two options for saving styles:
Save Settings As saves all the settings in Display Setup. The settings are given a name that is then listed in the Display Setup drop-down menu. Properly saved Setup Profiles are easily retrieved and applied to display windows. Click this button and enter the name you want to assign to the current settings.

Save To File saves current graphics display settings that you have defined. They are not as easily retrieved as when you use the Save Settings As option. This option could be used, as an example, when you are focused on preparing graphics for publications and wish to transfer formatting instructions from one display window to another. In the standard File Save dialog box, choose a file name for current styles and Standard Arrangement information.

Loading Styles

To apply a saved Display Profile to a Molecule Display window, choose a profile entry name from the list in the Display Setup drop-down menu or in the Settings Profile drop-down menu in the Display Setup dialog box. When you make the selection, Vector NTI loads the styles as well as other settings from the profile but does not apply them to the active window until you click the OK button. Modify the settings further, if you like; click the OK button to accept the new setup or click the Cancel button to leave the current window’s settings unchanged.

To retrieve styles saved in a separate setup file, press the Load From File button in the Graphics Display Settings box of the Display Setup dialog box. Choose a setup file from the standard File Open dialog box. Click the OK button in the Display Setup dialog box to apply new setup to the active Display window.

Saving Molecules

Vector NTI allows for two kinds of permanent storage for molecules:

- In the Vector NTI’s database.
  The database supports parent-descendant connections between molecules. It saves automatic tracking of modifications, search, and many other operations on the set of molecules as a whole.

- In a Vector NTI Molecule Document file.

Molecule Documents save all the information contained in a Molecule Display window such as graphical representation and molecule analysis results. In this format, the molecule itself is in a portable format suitable for information exchange between different Vector NTI users. Since the Vector NTI Molecule Document format is based on GenBank/GenPept format, other programs that are able to import GenBank/GenPept files can import Molecule Documents as well.

A Molecule Document for a constructed molecule does not contain or track the molecule’s parents and thus, unlike the database, it is not suitable for storing related molecules.
To save a molecule, click the **Save As** button or select **File > Save As**, opening the Save As dialog box (Figure 18.29):

![Figure 18.29 Save As dialog box](image)

Select the appropriate destination tab for the storage of your molecule.

**Save in <Molecule> Database**

Select the subset where the molecule is to be stored. Enter (or modify) the molecule name in the text box at the bottom of the dialog box and click **OK**.

**Save as (Molecule Document) File**

On the Save as File tab, browse for the destination where the document is to be saved. Enter a file name and click **OK**.

If Vector NTI is configured for Internet publishing, you will see a subsequent dialog box asking your permission to create an HTML file for the molecule. This HTML file will contain a brief description of the molecule and a hypertext link to the Molecule Document file.

**Creating A Molecule Shortcut File For A Database Molecule**

Molecule Shortcut files contain the name of a database molecule along with the Display Setup parameters. When you open a database molecule via the Shortcut file, Vector NTI uses the Display Setup parameters stored in the shortcut file.

To create a shortcut file, open the molecule in a Display window, set the desired Display Setup parameters and select **File > Create Shortcut**. In the standard File Save dialog box, enter a file name and location for the new Shortcut file.

**Molecule Export**

**Exporting a Molecule to a GenBank/GenPept, EMBL/SWISS-PROT or FASTA file**

Vector NTI allows you to easily export a molecule opened in a Molecule Display window to GenBank/GenPept, EMBL/SWISS-PROT, or FASTA file. For details about using Vector NTI’s complete export capabilities, refer to the Database Explorer, described in detail in Chapter 17.

To export a molecule to an external file, choose **File > Molecule Operations > Export**. Select the type and name of the file in the standard File Save dialog box.
Introduction

DNA/RNA and protein molecules can be edited in two environments:

- A Molecule Display window Graphics Pane, allowing editing of the molecule content in the graphics-oriented environment. Molecule Display windows give you immediate visual feedback and more automatic processing (that is, automatic adjustment of functional signal coordinates on sequence insertions and deletions).

- Molecule Editor, a text-based dialog box opened from a Molecule Display window of either molecule type.
  - Editing a molecule in the Molecule Editor is more convenient if you want, for example, to delete ten functional signals from the functional map.
  - Molecule Editor is the only way to create a molecule from scratch. (*Molecule Display windows cannot show molecules of zero length*).

Edit Operations in a Molecule Display Window

Molecule Editing

Vector NTI has two modes of editing operations in Graphics Panes.

- Molecule Editing mode (default) in which the Graphics Pane serves as a visual environment for viewing and editing a molecule map and sequence and for selecting DNA molecule fragments for construction and design. *In Molecule Editing mode, the molecule itself can be modified. Molecule Editing mode is the subject of this chapter.*

- Picture Editing mode where the *graphical display* of an image such as the format and arrangement of individual graphics objects is edited. This mode is useful for preparing publication-quality figures of the molecule. Picture Editing mode is discussed in Chapter 18 on page 226.
Note: The Vector NTI Database Explorer can be configured to launch independently or simultaneously with the Vector NTI workspace each time a Vector NTI session is initiated. From a Molecule Display window, select Edit > Options from the menu bar. On the General tab of the Options dialog box, clear the Open Local Explorer At Startup checkbox.

A Molecule Display window opens in Molecule Editing mode by default. Molecule Editing mode is used for viewing and editing graphical maps of a molecule, viewing ORFs and motifs (for DNA molecules), and editing molecule’s sequence. This mode is also used to select DNA molecule fragments for construction and design.

If the display window is in Picture Editing mode, the Edit Picture button ( ) is shown in depressed state and View > Edit Picture is checked.

Selection Techniques on the Graphical Map

The following techniques can be used for making selections in Molecule Editing mode:

- On the Graphics Map, point the cursor at a symbol or label. If pointing at a restriction site or motif, the cursor becomes a crosshair with a word “site” ( ). With the cursor paused at the site, a pop-up label appears, giving the type and site position. If pointing at a functional feature or an ORF, the cursor becomes a hand ( ) and upon pausing, it displays a message telling you the type, name, and location of the feature or ORF. Click on the symbol, site or its label to select it.

- Choose Edit > Set Selection. In the Set Selection dialog box, enter the coordinates of the region to be selected. Both start and end nucleotides are included in the selection, so to select the nucleotides or amino acids 50 through 60 you should enter From: 50, To: 60. Press OK. The coordinates of the selection are shown in the Selection Box on the status bar at the bottom of the screen.

- Click on the Set Selection box on the Status bar, opening the Set Selection dialog box. Enter the region coordinates in the Set Selection box. Click OK.

A selected region is marked with a selection wire-frame on the Graphics Pane (Figure 19.1). Concurrently the selection is highlighted in the Sequence Pane.

![Figure 19.1 Wire-frame selection on the Graphics Pane](image-url)
A caret (heavy black marker) appears at the “active” end on the wire-frame and the sequence at either the 5’ or 3’ end (a or c end for proteins). *The caret position is also shown in the caret position box on the status bar.*

The caret makes it easy to modify the ends of the selection in the Graphics Pane:

- To activate an end of the selection, click on it with the mouse or click on the HOME or END key.
- Hold the cursor over the active end; it changes shape to a cross-hair with the 5’ (↑), (3’, or a/c mark), click the mouse button and drag the end to the desired position (watch the selection box on the status bar while you are dragging it). Drag the 3’ end of the selection in the same way.
- To move the active end of the selection one residue at a time, hold the SHIFT + LEFT or RIGHT ARROW key to move the active end of selection one nucleotide/amino acid at a time.
- To move the caret 10 residues at a time, hold down the CTRL key (do not release the SHIFT key) with arrow keys. When the CTRL + SHIFT + LEFT OR RIGHT ARROW key.

**Note:** If you release your hold on the CTRL + SHIFT keys, you will lose your current selection.

**Setting The Caret Position On The Graphical Map**

To set the caret to a position of the molecule, choose *Edit > Set Caret Position* or click on the Set Caret box on the Status Bar. In the Set Caret Position dialog box, enter the coordinate of the nucleotide/amino acid following the caret. The coordinates start from 1; so to set the caret before the first nucleotide of the molecule you should enter 1. Click OK. The position of the caret appears on the Graphics and Sequence Panes and in the caret position box on the status bar.

**Note:** Setting a coordinate in a Set Selection or Set Caret Position dialog box replaces any previous selection.

**Selection Techniques for Choosing DNA Fragments for Recombination**

The simplest way to select DNA fragments for recombination is to use the Fragment Wizard, discussed in detail in Chapter 22 and Chapter 24. However, if you wish, you may select regions as described above.

To select more than one functional signal at a time, click on each while holding down the SHIFT key. The wire-frame will include all selected signals (plus intervening nucleotides). Press the TAB key to move the selection ends to the next gap between selected functional signals; press SHIFT + TAB moves the selection ends to the previous gap between selected functional signals. *If you prefer to select features without intervening nucleotides, select the fragments one by one and add them to the Goal List individually.*

To set a 3’ or 5’ end of the selection to a restriction site, press HOME or END to move the caret to the desired end of the selection and hold down the SHIFT key while clicking on the label of the restriction site. If an end of the selection zone is set to a restriction site, an @ symbol appears in front of that end’s coordinate in the status bar.

After the fragment is defined, click on the *Add Fragment to Goal List* button ( ), identify the role in construction or design the fragment will play, and press *Finish*.

**Editing the Molecule Map**

The elements of a molecule map can be divided into two categories.
1. Items explicitly described in the molecule data (features on DNA/RNA and protein molecules and restriction sites on un-sequenced regions of a DNA molecule). These CAN be edited.

2. Items calculated automatically from the DNA molecule’s nucleotide sequence (ORFs, motifs, and restriction sites on sequenced regions of the molecule) each time the molecule is displayed. These are not stored with the molecule; these CANNOT be edited.

Editing operations are as follows:

- **To edit a feature of the graphics map**, double-click on it or select it and choose **Edit > Feature Properties** from the shortcut menu or press ALT + ENTER. The Molecule Feature dialog box opens, where you can edit the properties of the selected object. The Molecule Feature dialog box for automatic objects does not contain editable fields.

- **To add a feature** to a molecule, select the corresponding region of the molecule and choose **Edit > New > Add Feature To FMap**. In the Molecule Feature dialog box, you can describe the new feature.

- **To add an ORF** to a DNA molecule’s Feature map, select the ORF you want to add and choose **Edit > New > Add ORF to FMap**. In the Molecule Feature dialog box, you can name the feature (all fields are initialized with the information taken from the selected ORF).

- **To add restriction sites** to unsequenced regions of a DNA molecule, move the caret to the desired position and choose **Edit > New > Add REN Site to RMap**. In the Restriction Sites dialog box, the Site Positions field shows the current caret position. You can enter more than one site position separated by space. Select the name of the enzyme and press **OK**. Vector NTI adds the sites modifying the nucleotide sequence accordingly.

- **To delete a feature** from the molecule map, select the feature. Choose **Edit > Delete Feature from FMap**, select **Delete Feature from FMap** from the shortcut menu or press DELETE on the keyboard. To remove “automatic” objects from the graphics map display, you must change the parameters in Display Setup.

**Notes about molecule editing:**

- If you have chosen an automatic feature to delete, a window appears reminding you that this feature cannot be deleted.

- You can rescind all molecule content edits by selecting **File > Revert to Saved**.

### Editing the Sequence from the Sequence Pane

Sequence editing operations in the Sequence Pane are similar to those in the Graphics Pane. Three editing operations make use of the standard Clipboard (Cut, Copy, and Paste) operations. Others allow you to insert literal fragments or modify existing fragments.

The copied sequence on the Clipboard is in the standard text format, for free exchange of the sequence data between Vector NTI and other applications.

To Paste (“insert”) a cut or copied sequence, set the cursor, which in the Sequence Pane becomes a vertical I-beam (I), to the insert position. Paste with a toolbar button or select **Edit > Paste Sequence**.

To insert a new sequence fragment manually, click the I-beam at the insertion position; choose the **Edit > New > Insert Sequence**. In the Insert Sequence dialog box, enter the sequence to be inserted. Click **OK**. If the inserted sequence will affect a feature, you are given the option to delete the feature (Delete or Delete All) or keep it (Keep or Keep All) in its modified form.

Once the sequence is inserted, you can delete it manually or select **File > Revert to Saved** Molecule in Database to return to the unedited version.
Edit Operations in Molecule Editor

This section explains the Vector NTI Sequence Editor, which can be used for creating new molecules from scratch (raw sequence), as well as for editing existing molecules.

To open the Molecule Editor dialog box:

From the Vector NTI Workspace/Molecule Display window:

- **File > Create New Sequence > Using Sequence Editor (DNA/RNA) or > Using Sequence Editor (Protein)** opens Molecule Editor for creating a new molecule from raw sequence.
- **Edit > <Molecule> Properties** opens Molecule Editor for the open molecule (Figure 19.2).

From Database Explorer:

- Select the DNA/RNA or Protein subset in the drop-down menu, select a molecule in Objects list and click on the **Edit** button ( ) or **<Molecule type >> Edit** on the Menu Bar.

The figure shows the DNA/RNA Molecule tab of the Molecule Editor opened from Database Explorer (Figure 19.2):

![Figure 19.2 DNA/RNA Molecule tab of the Molecule Editor](image)

The Molecule Editor contains information for one molecule at a time. Note these points about the Molecule Editor:

- Most tabs for protein or DNA/RNA are basically the same. Any variations are mentioned below.
- Molecule Editor opened from Database Explorer has one additional tab for sequence and feature editing. **Sequence editing of an open molecule must be performed in the Sequence Pane of the Molecule Display window itself.**
- To be edited with the Molecule Editor, a molecule must be fully processed. For instance, when you are trying to construct a DNA molecule from incompatible component fragments, Vector NTI saves the entered molecule’s data but does not process the molecule completely. It allows you to load molecule data from the database into the Construct/
Design Molecule dialog box, edit its component fragments, and reconstruct the molecule. But you cannot load a molecule into the Molecule Editor until it's fully processed.

- The Molecule Editor can be used to describe a new molecule from scratch. The Molecule Editor is almost identical for New Molecule (creating a new molecule from scratch) or for Edit Molecule. The New Molecule mode lets you enter all data including a sequence.

**Note:** You must enter a new molecule's nucleotide sequence using the Sequence Editor (Figure 19.6) to be able to save the new molecule. If the new DNA molecule is unsequenced, it is easier to construct it first from one DUMMY fragment using Construct/Design Molecule dialog box, and then enter its functional and restriction maps loading the molecule from the database into the Molecule Editor.

- If you have made any changes to a molecule in the Molecule Editor, you must press the Molecule Editor's OK button to enter the new information into the database.

**Note:** When a molecule is created during the design or construction process, its nucleotide sequence and functional map are generated automatically from its parents. You may add new functional signals to the map or delete or edit automatically-generated functional signals or edit an automatically generated nucleotide sequence. In that case, the system disconnects the molecule from its parents to avoid data inconsistency. The system prompts you if such a situation occurs.

Each of the tabs of DNA/RNA and protein Molecule Editors are discussed below.

### The General Information Tab

![Figure 19.3 The General Information tab displays general object data](image)

The General Information tab has the same format for all database objects (Figure 19.3). For existing molecules or other database objects, this tab allows you to only **review** object properties. For new objects **only**, you can enter a name for the new object in the active text box. Figure 19.3 displays the tab that opens with the Molecule Properties command for an existing molecule, thus there is no text box.
The DNA/RNA Molecule Tab

- **Molecule type**: Circular/linear and DNA/RNA
- **Extra-Chromosome Replication**: Transformation systems in which the molecule is capable of extra-chromosome replication.
- **Replicon Type**: The source of the replication origin for the molecule
- **Description**: Single-line description of the molecule.

The information in these fields appears in the General Description folder when the molecule is loaded into a Display window.

The **Protein tab** contains a one-line molecule description.

The Sequence and Maps Tab (DNA/RNA)

This tab appears only when the Edit <molecule> dialog box is opened from Database Explorer (Figure 19.5).
Editing Sequence

From the Sequence and Maps tab, click the **Edit Sequence** button to open the Sequence Editor to edit the sequence of the molecule (Figure 19.6).

![Figure 19.6 Using the Sequence Editor to edit the sequence of the molecule](image)

Edit the sequence using standard text editing techniques, Cut, Copy, and Paste through the Clipboard.

Click the **Find** button to specify a string to find in the sequence and the search direction (up or down). Click **Find Next** to launch the search and **Cancel** to leave the Find dialog box.

Click the **OK** button to save any changes you made in the sequence and to return to the Sequence and Maps tab of the Molecule Editor.

Editing the Feature Map

To add an item to a molecule’s Feature map, click the **Add** button in the Feature map pane of the Sequence and Maps tab. (This dialog box can also be accessed by: Select **Edit > New > Add Feature to FMap** or right-click on a feature in the Graphics Pane. Select **Feature Properties** from the shortcut menu.) The Molecule Feature dialog box appears (Figure 19.7):

![Figure 19.7 Molecule Feature dialog box](image)

**To add a feature**, select a particular Feature Type in the classification tree on the left. Enter the specific name of the feature in the Feature Name field. To see a list of specific names for each
feature type, press the **Browse** button ( ) next to the Feature Name field, opening the Feature Name dialog box, which lists the specific names available for that feature type (Figure 19.8):

![Feature Name dialog box](image)

*Figure 19.8 Feature Name dialog box, which lists the specific names available for that feature type*

After selecting a feature signal name, click **OK**.

Back in the Molecule Feature dialog box, select feature positions by using one of two formats:

- **Simple**: In the **From** and **To** fields, enter the starting and ending coordinates of the functional signal you are adding.  
  Check **Complementary** if the signal is located on the complementary strand.

  **Note**: Vector NTI uses the currently accepted convention for the coordinates of complementary features. All coordinates are given as if on the direct strand, from the leftmost to the rightmost nucleotide of a signal, as shown in Figure 19.9:

![Feature position coordinates](image)

*Figure 19.9 Defining feature position coordinates*

- **GenBank/EMBL format**: Set a complex location string that describes the positions of the exons and location of the features. *The full description of location string format can be found in the GenBank or EMBL sources*. For your convenience, you can always see the original location string of the feature.

  In the Description text box, enter notes about the feature you are adding.

  Click **OK** to add your feature and return to the Sequence and Maps tab.

**To edit a feature**: Double-click or highlight the feature in the Feature map pane and press **Edit**. The Molecule Feature dialog box (Figure 19.7) opens loaded with the feature you selected. Make any changes you want, and click **OK** to apply your changes or click **Cancel**.

**To remove a feature** from a molecule’s Feature map: Highlight the feature in the Feature map pane and click **Delete**.
**Editing the Restriction Map**

To add a restriction site to the molecule's restriction map: Click the **Add** button in the Restriction Map group of the Sequence and Maps tab (Figure 19.5), opening the Restriction Sites dialog box (Figure 19.10). (You can also access this dialog box by selecting **Edit > New > Add REN Site to RMap**.)

![Restriction Sites dialog box](image)

In the **Enzyme Name** field, enter the name of the restriction endonuclease or click the **Browse** button ( ) to see a list of RENs in the Vector NTI database.

In the **Site Positions** field, enter the nucleotide position numbers where the sites of the enzyme you have chosen are found, separated by spaces.

To edit a restriction site: Double-click it or highlight the item you want to edit and click **Edit**. The Restriction Sites dialog box appears, loaded with the site you selected. Make any changes you want. and click **OK** to register your changes or the **Cancel**. Vector NTI returns you to the Molecule Editor.

To remove an item from the restriction map: Highlight the item and click the **Delete** button in the Restriction Map group. *Sometimes Vector NTI does not allow removal of a restriction site if it is not entered manually.*

Sequence and Features Tab (Protein)

![Sequence and Features tab (Protein)](image)

This tab's functionality is the same as that of the Sequence and Maps tab of the DNA/RNA Molecule Editor, however, only sequence and Feature map editing are supported.
User Fields Tab

The User Fields tab (Figure 19.12), used for adding custom data for a molecule in the form of fields, has the same format for all database molecules. It contains a list of all User Fields currently applied to the object being edited as well as the list of blank fields defined in the local User Field Definition list but are not yet applied to that object.

- To fill or edit any field, double-click on the field name or select the field name and click the Change Value button. If the field you want to edit is not listed in the local User Field Definition list, Vector NTI asks you to enter the field's definition first. When the field definition is added to the local definition list, Vector NTI displays the appropriate field value dialog box.
- To clear any User Field for the object, select the field and click the Remove Value button.
- Refer to Chapter 17 for more user field details.

Comments Tab

The Comments tab (Figure 19.13) has the same format for all database objects. Enter text comments of unlimited length about the molecule.
Keywords Tab

The Keywords tab (Figure 19.14) has the same format for all database objects. This tab lets you enter keywords useful for database search.

To add a keyword for the molecule, type a new word or select an item in the list of existing keywords. Click the < Add button to move the keyword into the keyword list.

To remove an item from the keyword list, select item(s) click the > Remove button.

![Figure 19.14 Keywords tab](image)

Changing the Starting Coordinate of a DNA Molecule

You can change the starting coordinate of a molecule by choosing File > Molecule Operations > Advanced (DNA/RNA) > Change Starting Coordinate from the drop-down menu in a DNA/RNA Molecule’s Display window. Upon choosing this command, the Change Starting Coordinate dialog box appears, allowing you to designate the current molecule position that you want to define as the new starting coordinate:

![Figure 19.15 Change Starting Coordinate dialog box](image)

Click the OK button to accept the new start position. Vector NTI redraws the molecule, setting the New Start position as position 1. All features, restriction sites, etc. remain in the same positions relative to each other, but they are shifted in the display accordingly.

**Note:** When the starting coordinate of a molecule is changed, the change takes effect immediately. You are not asked to re-save the molecule to the database to register the change.

Making the Reverse Complement of a DNA Molecule

To make the reverse complement of a DNA/RNA molecule, open the molecule in a Molecule Viewing window in Vector NTI. With any pane activated, choose File > Make Reverse Complement from the drop-down menu. The Choose New DNA/RNA Molecule Name dialog box appears, allowing you to save the reverse complement.
Introduction

This chapter describes Vector NTI’s oligonucleotide analysis functions: the design, execution and analysis results for PCR primers, sequencing primers, and hybridization probes. The chapter also covers “silent mutation” analysis, ORFs and motif searches, calculation of restriction fragments, back translating and RFLP analysis.

The following table summarizes the various dialog boxes that open as you proceed through PCR, Sequencing and Hybridization Probe design:

<table>
<thead>
<tr>
<th>Analysis/Dialog Box</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find PCR Primers</td>
<td>Specify limits for PCR primer search such as length of target sequence, output options, attach restriction sites, etc.</td>
</tr>
<tr>
<td>Amplify Selection</td>
<td>Similar to Find PCR Primers except that primer hybridization domains upstream and downstream from the target sequence can be specified. Primers will be generated anywhere within the designated upstream and downstream domains.</td>
</tr>
<tr>
<td>Amplify Features</td>
<td>Similar to Find PCR Primers except that multiple features of a specific type can be selected for amplification. Use this dialog box when you need to amplify multiple features from a single molecule.</td>
</tr>
<tr>
<td>Long PCR</td>
<td>Similar to Find PCR Primers, but allows you to design primers for the amplification of regions that exceed 20 KB in length. Creates primers for a set of overlapped PCR products that cover the amplification region.</td>
</tr>
<tr>
<td>PCR Using Existing Oligos</td>
<td>Similar to Find PCR Primers, but, for the selected amplification region, allows you to search for suitable PCR primers from among those selected from the Vector NTI Oligo Database.</td>
</tr>
</tbody>
</table>

Table 20.1 Primer, oligo and probe dialog boxes
PCR Primers

Vector NTI locates primers for performing PCR analysis on a DNA molecule fragment selected in the Graphics Pane of a Molecule Display window. (Vector NTI can also evaluate user-defined primers.) After selecting the target sequence for primer design, the maximum and minimum product length and primer parameters are determined. Vector NTI evaluates, rates and sorts several primer options. You can further fine-tune the primers and annealing parameters if you wish, generate and display a PCR product in a Molecule Display window, save the primers (or product) in the database, order custom primers from Invitrogen, or use the product in recombinant cloning strategies.

Vector NTI provides seven different modes of PCR amplification:

- **Find PCR Primers** - Finds primers for a DNA selection within the selected region.
- **Amplify Selection** - Defines allowed regions upstream and downstream of the selected area for primer design.
- **Amplify Features** - Amplifies one or more selected features in a molecule.
- **Long PCR** - Designs primers for the amplification of regions that exceed 20 KB in length.
- **PCR Using Existing Oligos** - Searches for suitable PCR primers from among those selected from the Vector NTI Oligo Database.
- **Alignment PCR** - Design PCR primers for amplifying a region of aligned DNA molecules.
- **Multiplex PCR** - Designs PCR primers for amplifying several target sequence regions in one reaction.

This section covers preparation for PCR analysis, saving and loading PCR settings, the various PCR Primers dialog boxes, the results of PCR analysis presented in PCR Analysis folders, and operations with PCR primers and PCR products.
Open a Display window for the DNA molecule selected for PCR analysis. Select the target region of the molecule and choose Analyses > Primer Design > <PCR amplification method>, depending on the type of PCR analysis you want to perform. (These commands are disabled if an appropriate target is not defined.) This opens the corresponding PCR Analysis dialog box.

Saving and Loading PCR Settings

As you review the Find Primers, Amplify Selection, Amplify Features, Long PCR and PCR Using Existing Oligos dialog boxes, notice the Load and Save buttons on the lower left of each tab. These buttons allow you to save your PCR settings to a file and load the settings file in for subsequent analyses. This precludes having to reset the various parameters manually for frequently used PCR condition settings. You need only save/load the settings in one of the tabs and the parameters for all the tabs in the Find Primers dialog box will be saved/loaded.

**Important:** If you have made changes to the settings in a PCR Primers dialog box and want to save the settings to a file before running the analysis, you must first click the Apply button to register the changes before saving to a file. If the new settings are not applied, the preexisting settings are saved.

Find PCR Primers Dialog Box

The Find Primers dialog box (Figure 20.1) allows you to search for convenient sense and antisense primers according to your specifications and to check primers’ uniqueness. Clicking the More>> button on the Primer tab enables complete view of the Find Primers dialog box.
Primer Tab

The options in the Primer tab allow you to set parameters for PCR primer search and define primers and/or primer attachments, if desired:

### Primer Tab Parameters – Find Primers Dialog Box

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region of Analysis</td>
<td>Enter the start and end coordinates of molecule’s target region.</td>
</tr>
<tr>
<td>Product Length</td>
<td>Enter the maximum and minimum lengths of the molecule target region.</td>
</tr>
<tr>
<td></td>
<td>Note: Unless you specify differently here, the minimum amplicon length</td>
</tr>
<tr>
<td></td>
<td>may be less than the target sequence you selected.</td>
</tr>
<tr>
<td>Maximum Number of Output Options</td>
<td>Enter the number of sense-antisense primer pairs to be found. The actual</td>
</tr>
<tr>
<td></td>
<td>result may contain fewer than this number if there are not enough possible</td>
</tr>
<tr>
<td></td>
<td>primers.</td>
</tr>
<tr>
<td>Salt Concentration</td>
<td>Enter the PCR reaction salt concentration in mMol, if known.</td>
</tr>
<tr>
<td>Probe Concentration</td>
<td>Enter the value of probe concentration in pMol, if known.</td>
</tr>
<tr>
<td>dG Temperature</td>
<td>Enter the temperature in degrees Celsius to be used for calculating free</td>
</tr>
<tr>
<td></td>
<td>energy values.</td>
</tr>
<tr>
<td>Tm</td>
<td>Enter limits in degrees Celsius for primer melting temperature (T_m) (temp-</td>
</tr>
<tr>
<td></td>
<td>erature at which 50% of primer is a duplex) and the difference between T_m</td>
</tr>
<tr>
<td></td>
<td>for sense and antisense primers.</td>
</tr>
<tr>
<td>%GC</td>
<td>Enter the limits of G/C percentage in the primer and the difference between</td>
</tr>
<tr>
<td></td>
<td>GC percentages for sense and antisense primers.</td>
</tr>
<tr>
<td>Length</td>
<td>Enter primer length limits. Note: Nucleotide sequences such as RENs</td>
</tr>
<tr>
<td></td>
<td>attached to a primer’s 5’ end are included when calculating primer length.</td>
</tr>
<tr>
<td>DNA/RNA radio button</td>
<td>Select the type of target nucleotide sequence.</td>
</tr>
<tr>
<td>More&gt;&gt; and Less&lt;&lt; buttons</td>
<td>Extend or contract the lower portion of the Primers tab, which contains the</td>
</tr>
<tr>
<td></td>
<td>user-defined primer and primer attachment fields.</td>
</tr>
<tr>
<td>User-Defined Primers</td>
<td>Enter user-defined primer sequences or a primer from the oligo database.</td>
</tr>
<tr>
<td></td>
<td>The search engine checks the compatibility of the primers according to</td>
</tr>
<tr>
<td></td>
<td>primer parameters.</td>
</tr>
<tr>
<td>Attach to 5’ Terminus of (Anti)Sense Primer</td>
<td>Enter a short (≤&lt;18 bp) nucleotide sequence (if any) to be attached to the</td>
</tr>
<tr>
<td></td>
<td>5’ end of either primer, or choose from recognition sites of database RENs.</td>
</tr>
<tr>
<td></td>
<td>(To select RENs to add, click the Browse button ( ).) This sequence,</td>
</tr>
<tr>
<td></td>
<td>while considered in primer parameters, does not affect the calculation of</td>
</tr>
<tr>
<td></td>
<td>complementarity between primer and molecule. A sequence can be</td>
</tr>
<tr>
<td></td>
<td>attached to the primer whether or not the primers are user-defined or</td>
</tr>
<tr>
<td></td>
<td>designed by Vector NTI.</td>
</tr>
</tbody>
</table>

### Table 20.2 Primer tab parameters – Find Primers dialog box

**Note:** The calculation for Tm is dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Make sure to adjust these parameters according to your reaction conditions when performing your PCR analysis to ensure that you obtain accurate Tm values.
**Amplicon Tab**

Click the **Amplicon** tab (Figure 20.2). The options on this tab allow you to customize parameters relating to the resulting PCR product. %GC content for the product or a portion of the product and allowed bases adjacent to the primer annealing site can be specified.

Figure 20.2 Amplicon tab on Find Primers dialog box

<table>
<thead>
<tr>
<th>Amplicon Tab Parameters – Find Primers Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Amplicon %GC</strong></td>
</tr>
<tr>
<td><strong>Next to Primer Annealing Site</strong></td>
</tr>
</tbody>
</table>

Table 20.3 Amplicon tab parameters – Find Primers dialog box

**Structure Tab**

Click the **Structure** tab (Figure 20.3). The options on this tab set acceptable limits for nucleotide repeats, palindromes and hairpin loops for the primers. You can also check your primers/product for a selected group of restriction sites from this tab.

Figure 20.3 Structure tab on Find Primers dialog box
Pairs Tab

Click the **Pairs** tab (Figure 20.4). Options on this tab specify how closely parameters such as Tm and %GC, etc. must match between two primers in a generated primer set.

![Figure 20.4 Pairs tab on Find Primers dialog box](image)

<table>
<thead>
<tr>
<th><strong>Structure Tab Parameters – Find Primers Dialog Box</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nucleotide Repeats</strong></td>
</tr>
<tr>
<td><strong>Palindromes</strong></td>
</tr>
<tr>
<td><strong>Hairpin Loops</strong></td>
</tr>
<tr>
<td><strong>Check Hairpin Loops, Palindromes, Nucleotide Repeats and Dimers Only Within 3’ Region of ...</strong></td>
</tr>
<tr>
<td><strong>Check Primers For Restriction Sites From</strong></td>
</tr>
</tbody>
</table>

*Table 20.4 Structure tab parameters – Find Primers dialog box*

<table>
<thead>
<tr>
<th><strong>Pairs Tab Parameters – Find Primers Dialog Box</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tm Difference</strong></td>
</tr>
<tr>
<td><strong>%GC</strong></td>
</tr>
</tbody>
</table>

*Table 20.5 Pairs tab parameters – Find Primers dialog box*
Similarity Tab

Click the Similarity tab (Figure 20.5). The options on this tab determine the similarity relationship between the primers and the target sequence.

<table>
<thead>
<tr>
<th>Pairs Tab Parameters – Find Primers Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primer-Primer Complementarity</td>
</tr>
<tr>
<td>Primer-Primer 3’ End Complementarity</td>
</tr>
</tbody>
</table>

Table 20.5 Pairs tab parameters – Find Primers dialog box (continued)

Similarity Tab

Figure 20.5 Similarity tab on Find Primers dialog box

<table>
<thead>
<tr>
<th>Similarity Tab Parameters – Find Primers Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Best Fit</td>
</tr>
<tr>
<td>With Similarities &gt; = Similarity Threshold</td>
</tr>
<tr>
<td>Similarity Threshold</td>
</tr>
<tr>
<td>Last ... Nucleotides Must Have 100% Similarity</td>
</tr>
<tr>
<td>Similarity Between Ambiguous Nucleotides</td>
</tr>
</tbody>
</table>

Table 20.6 Similarity tab parameters – Find Primers dialog box
3’ End Tab

Click the 3’ end tab (Figure 20.6). The options on this tab allow you to set specifications for the 3’ end of the primers generated by Vector NTI. Parameters such as dG and specific nucleotide content for the 3’ end of both sense and antisense primers can be set here.

![Figure 20.6 3’ End tab on Find Primers dialog box](image)

<table>
<thead>
<tr>
<th>3’ End Tab Parameters – Find Primers Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dG &lt;=</strong></td>
</tr>
<tr>
<td><strong>Length for Analysis</strong></td>
</tr>
<tr>
<td><strong>Sense Primer 3’ Nucleotides</strong></td>
</tr>
<tr>
<td><strong>Antisense Primer 3’ Nucleotides</strong></td>
</tr>
</tbody>
</table>

Uniqueness Tab

Click the Uniqueness tab (Figure 20.7). The options on this tab determine the uniqueness of the primers generated with respect to the PCR product. These parameters can be used to help


ensure that generated primers bind to the desired template area with greater specificity than to the rest of the PCR product.

![Uniqueness tab on Find Primers dialog box](image)

**Figure 20.7  Uniqueness tab on Find Primers dialog box**

<table>
<thead>
<tr>
<th><strong>Uniqueness Tab Parameters – Find Primers Dialog Box</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Uniqueness Checks for</strong></td>
</tr>
<tr>
<td><strong>Max Allowed Similarity</strong></td>
</tr>
<tr>
<td><strong>Max Consecutive Match for Entire Primer</strong></td>
</tr>
<tr>
<td><strong>Primer 3’ End</strong></td>
</tr>
</tbody>
</table>

**Table 20.8  Uniqueness tab parameters – Find Primers dialog box**

**Qualities Tab**

Click the **Qualities** tab (Figure 20.8). These parameters govern primer quality by determining how much weight should be assigned parameters specified on the other tabs of the Find Prim-
ers dialog box. These values affect scoring functions that evaluate the quality rating of the primer sets generated.

![Figure 20.8 Qualities tab on Find Primers dialog box](image)

The importance factors are integers between 1 and 10 used in calculating the score evaluating primer/oligo quality. The lower the factor, the less weight given in the calculation. For example, for minimal importance, enter 1 in the appropriate box. For maximum importance, enter 10.

**Filters Tab**

Click the Filters tab (Figure 20.9). On this tab, you can select features that should either be excluded or included as regions to be considered for primer design. For example, if your sequence contains Repeat features, you may want to exclude them to obtain primers with greater specificity. Features can be added or removed by pressing the + and – buttons.

![Figure 20.9 Filters tab on Find Primers dialog box](image)

**Filters Tab Parameters – Find Primers Dialog Box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apply Feature Reject Filter</td>
<td>Check the Apply Feature Reject Filter box and press the + and – buttons to add or remove features from the list. This tells Vector NTI to avoid making primers to the region of the specified feature(s). A maximum allowed overlap of the primer with the feature’s 5’ and 3’ ends can be specified.</td>
</tr>
</tbody>
</table>

*Table 20.9 Filters tab parameters – Find Primers dialog box*
Amplify Selection Dialog Box

The Amplify Selection feature is similar to Find PCR Primers except that primer hybridization domains upstream and downstream from the target sequence can be specified. Amplify Selection picks primers to amplify the entire selection. If suitable primers cannot be found inside the selected region, the search will expand within the specified upstream and downstream flanking regions.

Like the Find PCR Primers dialog box, clicking the More>> button enables complete view of the dialog box. Only the tabs and parameters that differ from the Find PCR Primers dialog box are discussed here. For a description of the tabs and parameters common to the Amplify Selection and Find PCR Primers dialog boxes, see page 248 to page 254.

Primer Tab

The options in the Primer tab (Figure 20.10) allow you to set parameters for PCR primer search and define primers and/or primer attachments, if desired:

Amplify Features Dialog Box

Amplify Features is similar to Find PCR Primers except that Vector NTI will find the best primers for amplifying the designated feature(s) within the region of analysis indicated. Like the Find
PCR Primers and Amplify Selection dialog boxes, clicking the More>> button enables complete view.

Only the tabs and parameters that differ from the Find PCR Primers dialog box will be discussed here. For a description of the tabs and parameters common to the Amplify Features and Find PCR Primers dialog boxes, see page 248 to page 254.

**Primer Tab**

The options in the Primer tab (Figure 20.11) allow you to set parameters for PCR primer search and define attachments, if desired. User-defined primers are not allowed:

![Figure 20.11 Primer tab on Amplify Features dialog box](image)

**Primer Tab Parameters – Amplify Features Dialog Box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region of Analysis</td>
<td>Specify the 5’ and 3’ boundaries of the region for primer analysis. Features to be amplified must be within the chosen region. The number of features in the selection is indicated.</td>
</tr>
<tr>
<td>Every Amplicon Length</td>
<td>Specify the range for acceptable minimum and maximum product length. Vector NTI determines primers that will produce products within the set length range.</td>
</tr>
<tr>
<td>User-Defined Primers</td>
<td>This feature is NOT present.</td>
</tr>
</tbody>
</table>

*Table 20.11 Primer tab parameters – Amplify Features dialog box*
**Features Tab**

The **Features** tab (Figure 20.12) allows you to specify the features you want amplified, how many of the chosen features you want in one amplicon and whether primer-feature overlap is permitted:

![Features tab on Amplify Features dialog box](image)

**Table 20.12  Features tab parameters – Amplify Features dialog box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add Features to Amplify</td>
<td>Click the &lt;Add… button to display the feature list. From the list, choose the features you would like to amplify.</td>
</tr>
<tr>
<td>Features per Amplicon</td>
<td>Designate the minimum and maximum number of features you want to occur in the amplicons.</td>
</tr>
<tr>
<td>Feature-Primer Overlap</td>
<td>To allow feature-primer overlap for the selected features, check the appropriate box/boxes and specify the minimum and maximum permitted base overlap.</td>
</tr>
<tr>
<td>Find Primers in Feature</td>
<td>Specify whether VNTI should find primers either inside or outside of the amplified feature.</td>
</tr>
</tbody>
</table>

**Filter Tab**

The **Filter** tab (Figure 20.13) allows you to specify the features you do not want amplified as part of the product:

![Filter tab on Amplify Features dialog box](image)
Long PCR Dialog Box

Long PCR is similar to Find PCR Primers except that Vector NTI finds a set of primers for a group of overlapped PCR products that cover the selected amplification region. Applications for Long PCR include cloning of large genomic regions to search for polymorphisms and viral genome amplification.

**Note:** The selected region for amplification must exceed 20Kb in length for the Long PCR option to be enabled.

Like the Find PCR Primers dialog box, clicking the More>> button enables complete view.

Only the tabs and parameters that differ from the Find PCR Primers dialog box will be discussed here. For a description of the tabs and parameters common to the Long PCR and Find PCR Primers dialog boxes, see page 248 to page 254.

---

**Table 20.13 Filter tab parameters – Amplify Features dialog box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apply Feature Reject Filter</td>
<td>Check this box to enable the features reject filter. Add or remove features with the +, -, and Remove All buttons</td>
</tr>
<tr>
<td>Apply Feature Include Filter</td>
<td>Check this box to enable the features include filter. Add or remove features with the +, -, and Remove All buttons</td>
</tr>
<tr>
<td>5' End Overlap</td>
<td>Check the box and specify the maximum permitted overlap of the primer and the rejected feature(s) for the 5’ end.</td>
</tr>
<tr>
<td>3' End Overlap</td>
<td>Check the box and specify the maximum permitted overlap of the primer and the rejected feature(s) for the 3’ end.</td>
</tr>
<tr>
<td>Distance from Sense Primer 3’ End to Feature 5’ End</td>
<td>Check this box and specify the maximum distance in base pairs from the Sense Primer 3’ End</td>
</tr>
<tr>
<td>Distance from Antisense Primer 3’ End to Feature 5’ End</td>
<td>Check this box and specify the maximum distance in base pairs from the Antisense Primer 3’ End</td>
</tr>
</tbody>
</table>
Primer Tab

The options in the Primer tab (Figure 20.14) allow you to set parameters for Long PCR primer search and define attachments, if desired. User-defined primers are not allowed.

![Figure 20.14 Primer tab – Long PCR Analysis dialog box](image)

<table>
<thead>
<tr>
<th>Primer Tab Parameters – Long PCR Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplicon(s) Length</td>
</tr>
<tr>
<td>Overlap(s)</td>
</tr>
<tr>
<td>User-Defined Primers</td>
</tr>
</tbody>
</table>

Table 20.14 Primer tab parameters – Long PCR dialog box

PCR Using Existing Oligos Dialog Box

PCR Using Existing Oligos is similar to Find PCR Primers except that, for the selected amplification region, VNTI searches for suitable PCR primers from among those selected from the Vector NTI Oligo Database. Like the Find PCR Primers dialog box, clicking the More>> button enables complete view.

Only the tabs and parameters that differ from the Find PCR Primers dialog box will be discussed here. For a description of the tabs and parameters common to the PCR Using Existing Oligos and Find PCR Primers dialog boxes, see page 248 to page 254.
Primer Tab

The options in the Primer tab (Figure 20.15) allow you to set parameters for PCR primer search using primers from the VNTI Database and define attachments, if desired. User-defined primers are not allowed:

![Figure 20.15 Primer tab – PCR Using Existing Oligos dialog box](image)

**Primer Tab Parameters – PCR Using Existing Primers Dialog Box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense Primers</td>
<td>Click the <strong>Sense Primers</strong> button and select the desired primer(s) from the VNTI Oligo Database.</td>
</tr>
<tr>
<td>Antisense Primers</td>
<td>Click the <strong>Antisense Primers</strong> button and select the desired primer(s) from the VNTI Oligo Database.</td>
</tr>
<tr>
<td>User-Defined Primers</td>
<td>This feature is NOT present.</td>
</tr>
</tbody>
</table>

**Note:** Since you can choose the number of 3' and 5' primers, the PCR Using Existing Primers feature effectively enables you to analyze one 3' primer against an array of 5' primers or vice-versa.

Alignment PCR Analysis

The Alignment PCR feature in Vector NTI allows you to design PCR primers for amplifying a region of aligned DNA/RNA molecules. Using this feature you can design primer pair sets that will amplify any of the DNA/RNA molecules in the specified alignment. In this type of PCR analysis, the molecules are first aligned in AlignX and the alignment is subsequently analyzed using the Alignment PCR feature in Vector NTI.

**Basic Steps in the Alignment PCR Process**

1. **Defining DNA/RNA Molecules for Alignment PCR**
   In Vector NTI, you define the DNA/RNA molecules you want to use for Alignment PCR using the Alignment PCR tab of the Lists dialog box. From the Alignment PCR tab of the Lists dialog box, Align X is then launched so the alignment can be performed. Alternatively, you can drag and drop DNA/RNA molecules from the Vector NTI Database directly into AlignX or highlight molecules in...
the Database and choose **Align > AlignX – Align Selected Molecules** from the menu to open AlignX with the selected molecules, without using the Lists dialog box. However, using the Lists dialog box to initiate Alignment PCR gives you the additional capability of saving molecule lists for Alignment PCR that you do not have if you use AlignX directly.

**Note:** For more information about using AlignX, see Chapter 14 and Chapter 28.

To launch the Lists dialog box, choose **List > Show Lists** or from the Vector NTI Workspace/Molecule Viewing window and select the Alignment PCR tab or choose **List > Alignment PCR List** or **Analyses > Primer Design > Alignment PCR List** to display the Alignment PCR tab directly (Figure 20.16).

![Figure 20.16 Alignment PCR tab of the Lists dialog box](image)

**Note:** Initially the text box on the Alignment PCR tab of the Lists dialog box is empty.

- **Add** – Click the **Add** button to select molecules for Alignment PCR analysis from the Vector NTI DNA/RNA Database.
- **Load** – Click the **Load** button to load in a previously saved Alignment PCR molecule list file (.apcr file).
- **Save** – Click the **Save** button to save the current Alignment PCR molecule list to a file (.apcr file).
- **Delete** – Highlight a molecule on the list and click the **Delete** button to delete the molecule from the list.
- **Run** – Launches AlignX for the molecules on the Alignment PCR list.

Once you have configured the Alignment PCR molecule list, click the **Run** button to continue.

### Aligning DNA/RNA Molecules for Alignment PCR

The molecule alignment for Alignment PCR is performed in AlignX, as mentioned above. For more information on aligning molecules in AlignX, see Chapter 14 and Chapter 28.

### Performing the Alignment PCR Analysis

To launch the Alignment PCR dialog box, in the AlignX Alignment Pane highlight the region of the alignment you want to analyze, right-click and choose **Design PCR Primers** from the shortcut menu. The Alignment PCR dialog box displays (Figure 20.17), allowing you to configure the parameters for the PCR analysis.

**Note:** The Alignment PCR dialog box is part of Vector NTI and in the case where Vector NTI is not currently open, when the **Design PCR Primers** command is executed Vector NTI opens first, followed by the Alignment PCR dialog box.
Alignment PCR Dialog Box

Only the tabs and parameters of the Alignment PCR dialog box that differ from the Find PCR Primers dialog box will be discussed in this section. For a description of the tabs and parameters common to the Alignment PCR and Find PCR Primers dialog boxes, see page 248 to page 254.

Primer Tab

The options in the Primer tab (Figure 20.17) allow you to set parameters for the Alignment PCR primer search and define attachments, if desired. User-defined primers are not allowed.

![Figure 20.17 Primer tab on the Alignment PCR dialog box](image)

<table>
<thead>
<tr>
<th>Primer Tab Parameters – Alignment PCR Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primer/Every Molecule</strong></td>
</tr>
<tr>
<td><strong>Selection</strong></td>
</tr>
<tr>
<td><strong>Amplicon Length Min.</strong></td>
</tr>
<tr>
<td><strong>User-Defined Primers</strong></td>
</tr>
</tbody>
</table>

*Table 20.16 Primer tab parameters – Alignment PCR dialog box*
Results Tab

The options in the Results tab (Figure 20.18) allow you to specify the output format of the Alignment PCR primer search.

Figure 20.18 Results tab on the Alignment PCR dialog box

<table>
<thead>
<tr>
<th>Results Tab Parameters – Alignment PCR Dialog Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate Report</td>
</tr>
<tr>
<td>Save PCR Product(s) as DNA Molecule(s) to DB</td>
</tr>
<tr>
<td>Save Primer(s) as Oligo(s) to DB</td>
</tr>
</tbody>
</table>

Table 20.17 Results tab parameters – Alignment PCR dialog box

**Alignment PCR Analysis Report**

When the search is finished, if you have checked the Generate Report option on the Results tab, the Alignment PCR Report preview window displays the results of the analysis. The Alignment PCR Analysis Report dialog box appearance and function is similar to the Multiplex PCR Analysis report dialog box (Figure 20.29).

**Note:** Alignment PCR results must be saved to an external file; they cannot be saved in the Text Pane of a particular molecule (as is the case for Find Primers PCR results, for example), because they pertain to a group of molecules rather than an individual molecule.

**Multiplex PCR Analysis**

Multiplex PCR allows you to design PCR primers for amplifying several target sequence regions or features unambiguously in one reaction.
Basic Steps in the Multiplex PCR Process

The basic procedure for performing a Multiplex PCR analysis consists of defining the selections or features of the DNA/RNA molecules you want to use for Multiplex PCR analysis (selecting the region and adding it to the Lists dialog box) and performing the Multiplex PCR analysis on the selections and/or features you have designated.

Defining DNA/RNA Molecule Selections and/or Features

In Vector NTI you define the portion or feature(s) of a DNA/RNA molecule you want to use for Multiplex PCR by first opening the molecule in a Molecule Viewing window. Once the molecule is opened, select the region you are interested in analyzing or, in the case of Multiplex amplification of features, select a region containing the feature(s) you are interested in amplifying. After you have selected the region for analysis, launch the Lists dialog box and add the selection or feature type to the Multiplex PCR List tab.

To launch the Lists dialog box, choose List > Show Lists from the Vector NTI Workspace/Molecule Viewing window and select the Multiplex PCR tab or choose List > Multiplex PCR List or Analyses > Primer Design > Multiplex PCR List to display the Multiplex PCR tab directly (Figure 20.19). Initially the Multiplex PCR tab of the Lists dialog box is empty.

Note: Run GW and Run TOPO are tools for managing high throughput preparative PCR for these respective cloning technologies. They are described fully in Chapter 23.

![Figure 20.19 Multiplex PCR tab of the Lists dialog box with added molecule selections](image)

- **Add** – Click the Add button to add the highlighted region/feature(s) to the Lists dialog box. The Add Fragment to Multiplex PCR Analysis dialog box opens (Figure 20.20).
Important: If you have not highlighted a molecule or portion of a molecule or portion of a molecule in a Molecule Viewing window in Vector NTI, the Add button is greyed out and inaccessible.

![Figure 20.20 Add Fragment to Multiplex PCR Analysis dialog box](image1)

- **Add Whole Selection** – Adds the selected molecule region to the list for Multiplex PCR analysis (Figure 20.19)
- **Add Features Inside Selection** – From the molecule selection, adds the feature(s) that you designate to the list for Multiplex PCR analysis. Click the Add button (++) to launch the Add Feature dialog box (Figure 20.21) where you can choose the types of features you want to add.

![Figure 20.21 Multiplex PCR Add Features dialog box](image2)

- **Feature Type** – Highlight the feature type from the molecule selection that you want to add.
- **Add all features of this type** – Adds all features of the selected type within the molecule selection.
- **Add only features...with the following names** – Only adds features of the selected type with the names that you specify in the names list.

Once you have added the appropriate feature/feature names, click the OK button to accept the configuration. The Add Fragment to Multiplex PCR Analysis dialog box displays the new configuration (Figure 20.22):

![Figure 20.22 All CDSs in the molecule selection added](image3)
Note: Highlight a feature and click the Remove ( ) button to delete the feature from the Add Fragment to Multiplex PCR Analysis dialog box.

Click the OK button to add the fragment features to the Lists dialog box (Figure 20.23):

![Lists dialog box with added features](image)

- **Load** – Click the Load button to load in a previously saved Multiplex PCR molecule list file (.mpcr file).
- **Save** – Click the Save button to save the current Multiplex PCR molecule list to a file (.mpcr file).
- **Delete** – Highlight a molecule on the list and click the Delete button to delete the molecule from the list.
- **Run** – Launches the Multiplex PCR Analysis dialog box.

Performing the Multiplex PCR Analysis

Click the Run button on the Lists dialog box to launch the Multiplex PCR Analysis dialog box (Figure 20.25) where you can configure parameters for the PCR analysis (see **Multiplex PCR Dialog Box** below).

Click the OK button to begin the primers search. The Searching for PCR Primers dialog box displays, indicating the progress of the search (Figure 20.24):

![Searching for PCR Primers, Multiplex PCR](image)

**Multiplex PCR Dialog Box**

Only the tabs and parameters of the Multiplex PCR dialog box that differ from the Find PCR Primers dialog box will be discussed in this section. For a description of the tabs and parameters common to the Multiplex PCR and Find PCR Primers dialog boxes, see page 248 to page 254.
Primer Tab

The options in the **Primer** tab (Figure 20.25) allow you to set parameters for the Multiplex PCR primer search and define attachments, if desired. User-defined primers are not allowed.

![Primer Tab on the Multiplex PCR dialog box](image)

**Table 20.18  Primer tab parameters – Multiplex PCR dialog box**

<table>
<thead>
<tr>
<th><strong>Primer Tab Parameters – Multiplex PCR Dialog Box</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Find Primers in Selection/Feature</td>
<td>Set the parameter to search for primers outside or within the selection/feature.</td>
</tr>
<tr>
<td>Clustering – Multiplex Fold Minimum</td>
<td>Minimum number of selections/features with comparable parameters to be considered a cluster.</td>
</tr>
<tr>
<td>Clustering – Amplicon Length Difference Minimum</td>
<td>Minimum length in base pairs between the amplicons found for the different selections/features amplified within a cluster. A larger amplicon length difference helps distinguish amplified products when run on a gel, for example.</td>
</tr>
<tr>
<td>Clustering – Maximum/Minimum Clusters</td>
<td>Directs the program to divide the selections/features analyzed into more or less clusters. Minimizing the clusters in the case of large numbers of selections/features may lead to increased instances of random priming of found primers.</td>
</tr>
<tr>
<td>User-Defined Primers</td>
<td>This feature is NOT present.</td>
</tr>
</tbody>
</table>
Results Tab

The options in the Results tab (Figure 20.26) allow you to specify the output format of the Multiplex PCR primer search.

**Figure 20.26 Results tab on the Multiplex PCR Analysis dialog box**

<table>
<thead>
<tr>
<th><strong>Results Tab Parameters – Multiplex PCR Dialog Box</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Generate Report (see page 266)</strong></td>
<td>Displays a report of the Multiplex PCR results in a Preview window once the analysis is completed. You can print, copy or save the report to your hard drive from the Preview window (see Figure 20.29). <strong>Note:</strong> Multiplex PCR results must be saved to an external file; they cannot be saved in the Text Pane of a particular molecule (as is the case for Find Primers PCR results, for example), because they generally pertain to a group of molecules rather than an individual molecule.</td>
</tr>
<tr>
<td><strong>Save PCR Product(s) as DNA Molecule(s) to DB</strong></td>
<td>Allows you to save the resulting PCR products from clusters to the Vector NTI Database. Specify the name for the molecules and the DNA/RNA subset in which you want to save them.</td>
</tr>
<tr>
<td><strong>Save Primer(s) as Oligo(s) to DB</strong></td>
<td>Allows you to save the resulting PCR primers from clusters to the Vector NTI Database. Specify the name for the primers and the Oligo subset in which you want to save them.</td>
</tr>
</tbody>
</table>

**Table 20.19 Multiplex PCR dialog box, Results tab**

**Note:** At least one of the options on the Results tab MUST be selected before you are allowed to continue configuring parameters on the other tabs of the Multiplex PCR Analysis dialog box.
Cluster Tab

The options in the **Cluster** tab (Figure 20.27) specify how closely thermodynamic parameters must match between all primers generated within a cluster.

**Cluster Tab Parameters – Multiplex PCR Analysis Dialog Box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tm Difference</td>
<td>Enter difference in degrees Celsius between $T_m$ for primers in a cluster.</td>
</tr>
<tr>
<td>%GC</td>
<td>Enter the difference between GC percentages for primers in a cluster.</td>
</tr>
<tr>
<td>Primer-Primer Complementarity</td>
<td>Check the Permitted box for primer-primer complementarity; enter the minimum permitted value for duplex free energy.</td>
</tr>
<tr>
<td>Primer-Primer 3’ End Complementarity</td>
<td>Check the Permitted box for primer-primer 3’ end complementarity; enter the minimum permitted value for duplex free energy.</td>
</tr>
</tbody>
</table>

**Table 20.20 Cluster tab parameters, Multiplex PCR Analysis dialog box**

Uniqueness Tab

The options in the **Uniqueness** tab (Figure 20.28) determine the uniqueness of the primers generated with respect to the PCR product. These parameters can be used to help ensure that generated primers bind to the desired template area with greater specificity than to the rest of the PCR product.
Note: In the case of Multiplex PCR where many amplicons and primer sets may be used in the same reaction, often the conditions must be relaxed to allow some level of ambiguity in order to obtain results. Setting parameters to maximum stringency may result in no primers found.

![Uniqueness tab, Multiplex PCR Analysis dialog box](image)

**Uniqueness Tab Parameters – Multiplex PCR Dialog Box**

<table>
<thead>
<tr>
<th>Uniqueness Checks for</th>
<th>Choosing the <strong>Entire Molecule</strong> option for Multiplex PCR often results in the inability to form clusters. It is recommended that you choose the <strong>Only within Primers Annealing Sites in Amplicons</strong> option for Multiplex PCR.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Max Allowed Similarity</strong></td>
<td>75.0 %</td>
</tr>
<tr>
<td><strong>Max Consecutive Match for Entire Primer</strong></td>
<td>10 bp</td>
</tr>
<tr>
<td><strong>Max Consecutive Match for Primer 2 End</strong></td>
<td>7 bp Max Complementarity 50.0 %</td>
</tr>
<tr>
<td>10 bp</td>
<td></td>
</tr>
</tbody>
</table>

**Table 20.21 Uniqueness tab parameters, Multiplex PCR Analysis dialog box**

**Multiplex PCR Analysis Report**

When the search is finished, if you have checked the **Generate Report** option on the **Results** tab, the Multiplex PCR Report preview window displays the results of the analysis (Figure 20.29):

![Multiplex PCR Analysis results report](image)

**Figure 20.29 Multiplex PCR Analysis results report**

Three main folders in the report contain the details of the Multiplex PCR Analysis results:
Primer Design: PCR, Sequencing and Hybridization

Chapter 20

Clusterized (Amplicons in Clusters) – Contains the PCR results for selections/features amplified within each cluster, organized according to cluster number. Number of selections/features in the cluster appears in parenthesis in the cluster name. Clusters are numbered consecutively beginning with 1 and are listed according to number of selections/features in the cluster from highest to lowest as you move down the report.

Unclusterized (Amplicons not in any Clusters) – Contains the PCR results for selections/features amplified that could not be included as members of a cluster. The number of unclusterized products is shown in parenthesis after the folder name.

Unamplified – Contains the names of selections/features that could not be amplified using the parameters set in the Multiplex PCR dialog box. The number of unamplified products is shown in parenthesis after the folder name.

Click the appropriate buttons to print, copy and/or save the PCR report from the report preview dialog box.

Note: Multiplex PCR results must be saved to an external file; they cannot be saved in the Text Pane of a particular molecule (as is the case for Find Primers PCR results, for example), because they generally pertain to a group of molecules rather than an individual molecule.

The PCR Analysis Folder

After all of the parameters are defined in the PCR Analysis dialog box, click OK to begin analyzing/designing the primers.

During PCR primer analysis, Vector NTI generates a number of primer options that satisfy the conditions defined in the PCR Analysis dialog. These options are inserted into your Molecule Display window’s Text Pane as subfolders of the PCR Analysis folder (Figure 20.30).

Note: The PCR Analysis folder does not appear for Alignment PCR or Multiplex PCR Analyses. Alignment PCR and Multiplex PCR Analyses have their own reporting functions (see page 263 and page 270, respectively).

In the PCR Analysis Folder, the primer results are sorted in descending order according to their rating values calculated based on the importance factors assigned in the Qualities tab. The subfolder titles are numbered and show the length and rating of the PCR products they contain. The typical maximum rating is 171, although higher ratings are occasionally seen.

Each subfolder contains:
- The coordinates of molecule’s region included in the PCR product.
- Melting temperature for the product
- Optimum annealing temperature for the PCR reaction
- GC percentage of the product.

The next two folder sections describe the sense and antisense primers, each identified at the left with the primer icon ( ).

- The complementarity between the primer and its target sequence
- The actual primer sequence in the 5’ → 3’ direction (with attached sequence, if any)
- The primer length in nucleotides
- The primer melting temperature
- The primer’s GC percentage
- Primer values for enthalpy, entropy, and free energy.

If the **Check Primers for Restriction Sites From** box was checked on the Structure tab of the PCR Analysis dialog box (Figure 20.3), the primer section also contains a line showing restriction sites found in the primers and attached sequences. A number in parentheses after the cloning site name indicates the number of times that site occurs in the PCR product.

The last line in each product folder shows the difference in the melting temperatures and GC percentages of the sense and antisense primers.

**Operations with PCR Products**

**Note:** This section does not apply to Alignment PCR or Multiplex PCR Analyses. Alignment PCR and Multiplex PCR Analyses have their own PCR products operation functions (see page 263 and page 270, respectively).

Several operations can be performed on PCR products listed in the PCR Analysis folder in the Text Pane. Place the cursor on a PCR product folder line in the Text Pane and open the shortcut menu with its options:

- **Save as Molecule in Database** saves the product as an independent molecule in the database. This option first opens the New DNA/RNA Molecule tabbed dialog box (described in Chapter 19). Enter a name in the text box on the General tab. You can’t tab to the other pages without a name. The **Description** field on the DNA/RNA Molecule tab automatically defines the origin of the fragment. After entering data, click the **OK** button to save the product to the database.

- **Save to Database and Create Window** option saves the fragment in the database and immediately opens it in a display window. The option first opens the New Database Molecule dialog box, described immediately above. After entering information, click **OK**. The molecule is saved to the database and is immediately loaded into a Molecule Display window for your review.

- **Find PCR Product** option selects in the Graphics Pane and Sequence Panes the exact region of the molecule included in the PCR product. You can perform the same operation by clicking the **Find** button ( ) with the product folder selected or select **Edit > Find PCR product**.

**Operations with PCR Primers**

**Note:** This section does not apply to Alignment PCR or Multiplex PCR Analyses. Alignment PCR and Multiplex PCR Analyses have their own PCR primers operation functions (see page 263 and page 270, respectively).
Wherever you see the oligo symbol to the left of a primer sequence, ( ), you can perform various oligo functions. Place the mouse cursor on the oligo line and open a shortcut menu with these options:

- **Analyze** opens the Oligo Analysis dialog box (see Chapter 21) for performing oligo analysis.
- **Add to Oligo List** adds the primer to the Oligo List (see Chapter 21). This may be useful if you want to investigate possible duplexes between sense and antisense primers or order custom primers (For more information on ordering custom primers, see the next section.)
- **Save To Database** stores it in the database for future use.

**Ordering Custom Primers**

In Vector NTI Advance 10, you can order custom primers that you have generated and analyzed in the software. Custom primers are synthetic oligonucleotides made with your specified sequence for use in a variety of applications from PCR and sequencing to probes for gene detection.

To initiate the order process, select the oligo from the PCR Analysis Text Pane folder by placing the mouse cursor on the oligo line (indicated by the oligo symbol ( ) to the left of a primer sequence) and open a shortcut menu with a right click. Choose the **Add to Oligo List** feature. In the New Oligo dialog box, enter the name of your oligo and click **OK**. (Additional information pertaining to the New Oligo, such as user fields and keywords, can be entered in the New Oligo dialog box. For more information on adding a new oligo to the database see page 72.) Vector NTI verifies the oligo has been added.

Open the Oligo List dialog box by selecting **List > Oligo List**. Select the appropriate oligo(s) and click the **Order** button ( ). The Invitrogen webpage opens automatically, and your primer sequence is entered in the custom primer order form. Enter your name and any custom modification you want, such as standard deoxynucleotides, modified bases, 5´ modified nucleotides (including fluorescent dyes, enzyme conjugates), and S-oligos for antisense studies. Specifications also include scale of synthesis, level of purity and 5´ or 3´ modifications. Press the **Add to Cart** button and proceed to checkout, following the steps outlined on the website.

**Saving, Viewing and Printing PCR Analysis Results**

**Note:** This section does not apply to Alignment PCR or Multiplex PCR Analyses. Alignment PCR and Multiplex PCR Analyses have their own PCR saving/reporting functions (see page 263 and page 270, respectively).

The Vector NTI Advance data type, **Analysis Results**, enables you to save PCR analysis results to the Vector NTI Advance Database. Saving PCR analysis results to the Database, allows you to view the results in a Molecule Viewing window at a later date and print the results, including the PCR Settings used for the analysis.

**Saving PCR Analysis Results to the Database**

To save PCR analysis results to the Database, after performing the analysis, right-click on the PCR Analysis folder in the Text Pane of the Molecule Viewing window and choose **Save As**
**Analysis Result** from the shortcut menu. In the dialog box that opens, name the PCR analysis and select the Analysis Results subset to which you want to save it (Figure 20.31):

![Choosing new analysis result name](image)

*Figure 20.31 Saving PCR results to the Database*

**Viewing Saved PCR Analysis Results**

To view PCR analysis results saved in the Database, open the molecule for which the analysis was performed and saved. In Display Setup, make sure that **Analyses** is checked and that the Analyses Setup is configured appropriately (see Analyses Setup, page 221). This tells Vector NTI to display the designated analyses in the Molecule Viewing window as part of your display profile (Figure 20.32):

![Saved PCR results displayed in Molecule Viewing window](image)

*Figure 20.32 Saved PCR results displayed in the Molecule Viewing window*

**Tip:** You may want to save one display profile with **Analyses** turned on and one with **Analyses** turned off, so that you may easily choose either profile from your list of saved profiles.

**Printing PCR Analysis Results and Settings**

When the Analyses parameter is turned on in the display profile, Analysis Results are shown in the Graphics and Sequence Panes and the details of the analysis appear in the Text Pane in the External Analyses folder in a subfolder named according to the name you specified when you saved the result to the Database. The analysis settings you used to perform the PCR analysis appear in the PCR Analysis Settings subfolder.

You can copy with the Camera the results of your PCR analysis, including PCR settings, and print them as follows:

1. Open the folders of interest in the Text Pane.
2. Highlight the desired information.
3. Copy the selection to the Clipboard (or file) using the Camera tool.
4. Paste from the Clipboard (or open the saved file) into a text editor program.
5. Print from the text editor program.

**Sequencing Primers**

To find primers for sequencing a DNA molecule fragment when the molecule is opened in a DNA Molecule Display window, select the target sequence of the molecule—the region to be sequenced. Primers will be generated anywhere within the designated the “Primer Hybridizing Domain” (upstream and downstream). If the sequencing region is long enough, it is divided by Vector NTI into sequencing domains, areas in which a single sequencing reaction will take place. The size of the primer hybridizing domain may then be set (the region in which primers are sought) as well as other desired primer parameters. Several primer options are evaluated and sorted from best to worst.

To prepare for sequencing primer design, open a sequence or Molecule Display window for the molecule and select the region for analysis, using ordinary selection techniques (described in Chapter 3).

Select **Analyses > Primer Design > Sequencing Primers**. *This command is disabled if a selection zone is not defined.* This opens the Sequencing Primers dialog box (Figure 20.33).

**Sequencing Primers Dialog Box**

The options in this dialog box allow you to set parameters for sequencing primer search:

<table>
<thead>
<tr>
<th><strong>Sequencing Primers Dialog Box Parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA/RNA</td>
</tr>
<tr>
<td>Sequencing Region</td>
</tr>
<tr>
<td>Sequencing Domain</td>
</tr>
</tbody>
</table>

*Table 20.22 Sequencing Primers dialog box parameters*
After all parameters are defined, click OK to search for the primers. During primer analysis, Vector NTI generates a number of primer options that satisfy the conditions you have defined. The primer options appear in the Text Pane as subfolders (Figure 20.30).

**Primer Similarity Dialog Box**

Click the **Primer Similarity** button in the [Sequencing Primer] [Hybridization Probe] Analysis box to open this dialog box to specify similarity requirements for the primer(s)/probe (Figure 20.34). (See Table 20.23)

<table>
<thead>
<tr>
<th><strong>Primer Parameters</strong></th>
<th>Click the <strong>Primer Parameters</strong> button to set parameters for sequencing primers in the Primer/Oligo Parameters dialog box (Figure 20.35). (See Table 20.24 for description.)</th>
</tr>
</thead>
</table>

**Primer Quality Specifics**

Click the **Primer Quality Specifics** button to open the Primer/Oligo Quality Specifics dialog box (Figure 20.36) for assigning “importance factors” to sequencing primers.

<table>
<thead>
<tr>
<th><strong>Sequencing Primers Dialog Box Parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Primer Hybridizing Domain</strong></td>
</tr>
<tr>
<td><strong>Maximum Number of Primer Options for Each Domain</strong></td>
</tr>
<tr>
<td><strong>User-Defined First Primer</strong></td>
</tr>
<tr>
<td><strong>Analyze</strong></td>
</tr>
<tr>
<td><strong>Primer Similarity</strong></td>
</tr>
<tr>
<td><strong>Primer Parameters</strong></td>
</tr>
<tr>
<td><strong>Primer Quality Specifics</strong></td>
</tr>
</tbody>
</table>

**Table 20.22  Sequencing Primers dialog box parameters (continued)**

After all parameters are defined, click OK to search for the primers. During primer analysis, Vector NTI generates a number of primer options that satisfy the conditions you have defined. The primer options appear in the Text Pane as subfolders (Figure 20.30).

<table>
<thead>
<tr>
<th><strong>Primer Similarity Dialog Box Parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best Fit</strong></td>
</tr>
<tr>
<td><strong>&gt; = Similarity Threshold</strong></td>
</tr>
<tr>
<td><strong>Similarity Threshold</strong></td>
</tr>
</tbody>
</table>

**Table 20.23  Primer Similarity dialog box parameters**
Primer/Oligo Parameters Dialog Box

Click the **Primer Parameters** button in the [Sequencing Primer] [Hybridization Probe] Analysis box to open this dialog box to specify acceptable limits for the interaction of the primer(s) or hybridization probes (oligo) with the selected target sequence (Figure 20.35).

### Table 20.24 Primer/Oligo parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dG Temperature</td>
<td>Enter the temperature in degrees Celsius to be used for calculating free energy values.</td>
</tr>
<tr>
<td>Probe Concentration</td>
<td>Enter the value of probe concentration in pMol, if known.</td>
</tr>
<tr>
<td>Salt Concentration</td>
<td>Enter the value of salt concentration in mMol, if known.</td>
</tr>
<tr>
<td>% Formamide</td>
<td>When hybridizing in formamide, enter the formamide concentration in %.</td>
</tr>
<tr>
<td>Length</td>
<td>Enter primer length limits. Note: Nucleotide sequences such as RENs attached to a primer’s 5’ end are included when calculating primer length.</td>
</tr>
<tr>
<td>%GC</td>
<td>Enter the limits of G/C percentage in the primer and the difference between GC percentages for sense and antisense primers.</td>
</tr>
<tr>
<td>Tm</td>
<td>Enter limits in degrees Celsius for primer melting temperature (T_m) (temperature at which 50% of primer is a duplex) and the difference between T_m for sense and antisense primers.</td>
</tr>
</tbody>
</table>
Table 20.24 Primer/Oligo parameters (continued)

Primer/Oligo Parameters

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3' End</td>
<td>Specify parameters for the primer’s 3’ end. In the Length for Analysis field, enter the length of the primer 3’ region that should be analyzed. In the ( dG ) ( \leq ) field, specify the maximum permitted value of 3’ end free energy. Check the nucleotide boxes to specify permitted last primer nucleotides.</td>
</tr>
<tr>
<td>Hairpin Loop Stem Length ( \geq )</td>
<td>Enter the minimum number of base pairs in a hairpin stem. (This value is also used as a minimum stacking length for primer-primer complementarity and primer-primer 3’ end complementarity.)</td>
</tr>
<tr>
<td>Palindromes ( \leq )</td>
<td>Enter the maximum permitted length of palindromes in primers.</td>
</tr>
<tr>
<td>Nucleotide Repeats ( \leq )</td>
<td>Enter the maximum permitted length of nucleotide repeats in primers.</td>
</tr>
<tr>
<td>Hairpin Loops</td>
<td>Check the Permitted box for hairpin loops; enter the minimum permitted value for free energy of hairpin loops. Primers with hairpin loops which have free energy values ( \leq ) to this number will be accepted.</td>
</tr>
<tr>
<td>Primer-Primer Complementarity</td>
<td>Check the Permitted box for primer-primer complementarity; enter the minimum permitted value for duplex free energy.</td>
</tr>
<tr>
<td>Primer-Primer 3’ End Complementarity</td>
<td>Check the Permitted box for primer-primer 3’ end complementarity; enter the minimum permitted value for duplex free energy.</td>
</tr>
<tr>
<td>Check All Features Only Within 3’ Region of …</td>
<td>Check this box and enter the length of a 3’ region if all of a primer’s features (repeats, palindromes, hairpin loops, dimers) should be checked only within that 3’ region. (If this box is empty, the whole primer will be evaluated.)</td>
</tr>
<tr>
<td>Primers/Oligos Must Be Unique With Similarity …</td>
<td>Check this box and enter the similarity threshold to check primer uniqueness on the molecule. Primers which have parasitic hybridization with similarity ( \leq ) this threshold will be rejected. Note: this similarity threshold must be ( \leq ) the minimum similarity required for hybridization of user-defined primers (if any).</td>
</tr>
</tbody>
</table>

Primer/Oligo Quality Specifics Dialog Box

Click the [Primer]/[Oligo] Quality Specifics button in the [Sequencing Primer] [Hybridization Probe] Analysis box to open this dialog box to assign “importance factors” to different primer/oligo parameters (Figure 20.36).

![Figure 20.36 Primer/Oligo Quality Specifics dialog box](image)

The importance factors are integers between 1 and 10 used in calculating the score evaluating primer/oligo quality. The lower the factor, the less weight given in the calculation. For example, for minimal importance, enter 1 in the appropriate box. For maximum importance, enter 10.
The Sequencing Primers Folder

The Sequencing Primers folder (Figure 20.37) contains one subfolder with primers for each sequencing domain in the analyzed region. Subfolder titles are numbered and show the length and boundaries of the sequencing domains. Each subfolder contains:

- The primer sequence, in the 5' → 3' direction
- The primer hybridization site and primer similarity for that site
- The primer length in nucleotides
- Melting temperature for the primer
- GC percentage of the primer
- Values for enthalpy, entropy, and free energy

Wherever you see the oligo symbol to the left of a primer sequence, ( ), you can perform various oligo functions. The shortcut menu associated with these primers, launched by right clicking on the primer sequence, has these options:

- **Analyze** opens the Oligo Analysis dialog box (see Chapter 21) to perform oligo analysis.
- **Add to Oligo List** adds the primer to the Oligo List (see Chapter 21). This can be useful to investigate possible duplexes between sense and antisense primers or to order custom primers. For more information on ordering custom primers, see page 273.
- **Save To Database** stores it in the database for future use.

Hybridization Probes

Hybridization probes analysis is used to design oligonucleotides that will hybridize with a selected molecule fragment, within specified parameters. Vector NTI can generate a set of oligos or use user-defined or database-stored oligos to test for hybridization efficiency with a target molecule.

Open a Molecule Display window for the molecule and select the region for analysis, using ordinary selection techniques (described in Chapter 3).

Select **Analyses > Primer Design > Hybridization Probes**. This command is disabled if a selection zone is not defined. This opens the Hybridization Probes dialog box (Figure 20.38).
Hybridization Probes Dialog Box

Set parameters in this dialog box for successful hybridization of your probes with your target fragment:

<table>
<thead>
<tr>
<th>Hybridization Probes Dialog Box Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search Region</strong></td>
</tr>
<tr>
<td><strong>DNA/RNA</strong></td>
</tr>
<tr>
<td><strong>Maximum Number of Output Options</strong></td>
</tr>
<tr>
<td><strong>User-Defined Oligo(s)</strong></td>
</tr>
<tr>
<td><strong>Analyze</strong></td>
</tr>
<tr>
<td><strong>Oligo Similarity</strong></td>
</tr>
<tr>
<td><strong>Oligo Parameters</strong></td>
</tr>
<tr>
<td><strong>Oligo Quality Specifics</strong></td>
</tr>
</tbody>
</table>

Table 20.25 Hybridization Probes dialog box parameters
The Hybridization Probes Folder

Hybridization probe analysis generates oligos that will hybridize with the target sequence and stores them in the Hybridization Probes folder (Figure 20.39) in the Text Pane.

![Hybridization Probes: 143 - 471](image)

Figure 20.39 The Hybridization Probes folder

If a user-defined oligo was specified, the Hybridization Folder contains only that oligo, if it could hybridize in the defined conditions. The Hybridization folder contains one subfolder for each probe for the analyzed region.

Subfolder titles show the name of the oligo contained within.

For each oligo, the subfolders contain:

- The nucleotide sequence of the probe in the 5'-3' direction
- The oligo hybridization site and similarity
- The oligo length in nucleotides
- The oligo melting temperature
- The oligo GC percentage
- Oligo values for enthalpy, entropy, and free energy.

Wherever you see the oligo symbol to the left of a probe sequence, ( ), you can perform various oligo functions. Place the mouse cursor on the oligo line and right click to open a shortcut menu with these options:

- **Analyze** opens the Oligo Analysis dialog box (see Chapter 21) to perform oligo analysis.
- **Add to Oligo List** adds the primer to the Oligo List (see Chapter 21). This may be useful if you want to investigate possible duplexes between sense and antisense primers or order custom primers. For more information on ordering custom primers, see page 273.
- **Save To Database** stores it in the database for future use.
Introduction

This chapter describes Vector NTI’s search and analysis functions, including Oligo analyses, ordering Custom Primers, Motifs and ORF searches, Mutagenesis, Translation, Back Translation and various Restriction Enzyme analyses.

Oligonucleotide Analyses

This section describes Vector NTI’s oligonucleotide functions, including the Oligo List and the Oligo Editor, Thermodynamic Properties analysis, and Oligo Duplexes dialog boxes. Vector NTI Oligo analysis tools are summarized in the following table:

<table>
<thead>
<tr>
<th>Analysis/Dialog Box</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo List</td>
<td>Stores a list of oligos for current use. From this dialog box, add new, edit, analyze, or order oligos.</td>
</tr>
<tr>
<td>Edit/New Oligo</td>
<td>Enter new or edit oligo data, including name, sequence, user fields, comments, keywords.</td>
</tr>
<tr>
<td>Thermodynamic Properties (Oligo Analysis)</td>
<td>Set limits for annealing conditions with target sequence; analyze oligo for meeting specified criteria.</td>
</tr>
<tr>
<td>Dimers and Hairpin Loops</td>
<td>Examine oligo’s dimer and hairpin loop structures.</td>
</tr>
<tr>
<td>Oligo Duplexes</td>
<td>Generate all possible duplexes of selected oligos.</td>
</tr>
</tbody>
</table>

Table 21.1 Oligo Analysis Tools

The Oligo List

The Oligo List is a data clipboard for listing and analyzing oligonucleotides from a display window or from the database. Oligonucleotides can be user-defined or determined by selecting a target sequence and letting VNTI design an oligo for the sequence. From this central list, you can add new oligos or select existing ones for editing and analysis.
To open the Oligo List tab of the Lists dialog box (Figure 21.1), in the Vector NTI workspace or Molecule Viewing windows, click the Open Oligo List button ( ) on the Main Toolbar or choose List > Oligo List from the menu.

**Figure 21.1 Oligo List dialog box**

**Note:** Initially, the Oligo List is empty.

Button options in the dialog box are summarized as follows:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
</table>
| Add                | Opens the Oligo Editor dialog box (Figure 21.2) to enter a new oligo to the Oligo list.  
  **Note:** The calculation does not include an H atom at the 5’ end and an OH group at the 3’ end of the oligo. It does include the PO4 group at the 5’ end. |
| Load               | Loads oligos into the list from the VNTI oligo database.                   |
| Save               | Click **Save to Database** button ( ) to save it in the database.          |
| Edit               | Opens the Oligo Editor dialog box (Figure 21.2) to edit the selected oligo. |
| Remove             | Removes the selected oligo from the Oligo List.                           |
| Analyze            | Initiates oligo analysis on the selected oligo. (See Figure 21.1)          |
| Duplexes           | Investigates the duplexes on one or more oligos. (See Figure 21.1)         |
| Order              | Link to Invitrogen to order custom primers (see section on Ordering Custom Primers below). |

**Table 21.2 Oligo List dialog box parameters**

**Order Custom Primers**

In Vector NTI Advance 10, you can order custom primers that you have generated and analyzed in the software. Custom primers are synthetic oligonucleotides made with your specified sequence for use in a variety of applications from PCR and sequencing to probes for gene detection.

Open the Oligo List dialog box by selecting **List > Oligo List** (Figure 21.1). Select the appropriate oligo(s) and click the **Order** button ( ). The Invitrogen webpage opens automatically, and your primer sequence is entered in the custom primer order form. Enter your name and any custom modification you want, such as standard deoxynucleotides, modified bases, 5’ modified nucleotides (including fluorescent dyes, enzyme conjugates), and S-oligos for antisense studies. Specifications also include scale of synthesis, level of purity and 5’ or 3’ modifications. Press the **Add to Cart** button and proceed to checkout, following the steps outlined on the website.
Oligo Editor Dialog Box

To open the New/Edit Oligo dialog box, called the Oligo Editor:

- **Database Explorer**: select the Oligo subset in the drop-down menu, select an oligo from the list in the Object Pane and double-click on it, select **Oligo > Edit** or press the **Edit** button ( ).

- **Molecule Viewing window**: In the Text Pane showing oligos, highlight an oligo and either right click and select **Add to Oligo List** from the shortcut menu or click on the **Add to Oligo List** button ( ), or click **Add** in the Oligo List dialog box (see previous section).

The dialog box that opens is the same for a new oligo or one to be edited (Figure 21.2).

![Figure 21.2 Oligo Editor dialog box](image)

The Oligo Editor contains information for one oligo at a time. In the New Oligo dialog box, you will enter all data including the oligo sequence (except when opened from an oligo in a Text Pane folder. In that case, the sequence is already entered.) To enter or modify information about an oligo on the tabs:

**General**: Enter or edit the name of the oligo.

**Oligo**:

- Select a radio button for the oligo type.
- Check the **Reverse Complementary** box to replace the oligo with its complementary sequence.
- In the Nucleotide Sequence box, enter or edit the standard I.U.B. base code characters for the molecule sequence. (See Appendix C.)
- Enter the oligo’s description.

**User Fields**: Modify custom data in a form of fields (see Chapter 17 for User Fields details.)

**Comments**: Enter text comments about the oligo.

**Keywords**: To add a keyword for the oligo, type a new word or select an item in the list of existing keywords. Click the **Add** button to move the keyword into the oligo’s keyword list. To remove an item from the keyword list, select it and press the **Remove** button.

Exit Oligo Editor by clicking **OK** or **Cancel**.
Thermodynamic Properties Analysis (Oligo Analysis) Dialog Box

Oligonucleotides for analysis can be selected by selecting a target sequence and letting VNTI design an oligo for the sequence or entering a user-defined oligo.

- Select a target sequence in an open Molecule Display window, and select Analyses > Oligo Analyses > Thermodynamic Properties; the oligo selected by Vector NTI is displayed in the text box of the Oligo Analysis dialog box (Figure 21.3) that opens.

or

- Open the Oligo Analysis dialog box first with Analyses > Oligo Analyses > Thermodynamic Properties, and then enter a user-defined oligo sequence in the Oligonucleotide text box.

You can also select an oligo listed in a Text Pane folder and select Analyze from the shortcut menu.

![Figure 21.3 Oligo Analysis dialog box](image)

Select radio buttons for molecule type, and check Complementary to replace the oligo with its complement.

The Parameters column lets you accept default settings or specify acceptable limits for the interaction of the oligo with its selected target sequence. Settings are summarized as follows:

<table>
<thead>
<tr>
<th>Oligo Analysis Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>dG Temperature (°C)</td>
<td>Enter the temperature in degrees Celsius to be used for calculating free energy values.</td>
</tr>
<tr>
<td>Probe Conc. (pMol)</td>
<td>Enter the value of probe concentration in pMol.</td>
</tr>
<tr>
<td>Salt Conc. (mMol)</td>
<td>Enter the value of salt concentration in mMol.</td>
</tr>
<tr>
<td>% Formamide</td>
<td>Enter the formamide concentration in %.</td>
</tr>
<tr>
<td>3' End Length (bp)</td>
<td>Specify the length of the oligo 3’ region that should be analyzed.</td>
</tr>
<tr>
<td>Palindromes</td>
<td>Enter the minimum length of palindromes to search for.</td>
</tr>
<tr>
<td>Nucleotide Repeats (bp)</td>
<td>Enter the minimum length of nucleotide repeats to search for.</td>
</tr>
</tbody>
</table>

Table 21.3 Oligo Analysis parameters
Note: The calculation for Tm is dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Make sure to adjust these parameters according to your reaction conditions when performing your oligo analysis to ensure that you obtain accurate Tm values.

After these parameters are set, click Analyze. Results are summarized in the same dialog box as follows:

<table>
<thead>
<tr>
<th>Oligo Analysis Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mol. Wt</td>
</tr>
<tr>
<td>%GC</td>
</tr>
<tr>
<td>Therm. Tm</td>
</tr>
<tr>
<td>% GC Tm</td>
</tr>
<tr>
<td>dG</td>
</tr>
<tr>
<td>3' End dG</td>
</tr>
<tr>
<td>dH</td>
</tr>
<tr>
<td>dS</td>
</tr>
<tr>
<td>Palindromes</td>
</tr>
<tr>
<td>Nucleotide Repeats</td>
</tr>
</tbody>
</table>

Click the Dimers & Hairpin Loops button to examine the oligo’s dimers and hairpin loops. Both are undesirable secondary structures that interfere with hybridization or amplification efficiency.
**Dimers & Hairpin Loops Dialog Box**

![Image of Dimers & Hairpin Loops dialog box]

*Figure 21.4 Dimers & Hairpin Loops dialog box*

**Note:** In the graphical depiction of dimers and hairpin loops, vertical lines indicate the primary interaction, based on the stem length set and plus symbols indicate secondary interactions. The greater the dG value, the weaker the interaction; secondary interactions are not considered in the dG calculation.

**Dimers**

Dimers are formed when one oligo anneals to another oligo. See the Dimer example in the upper pane of Figure 21.4. Examine the dimers the oligo contains. Click the Next button (>>) to view the next dimer. Click the Previous button (<<) to view the previous dimer.

**Hairpin Loops**

Hairpin loops are formed when an oligo doubles back on itself. See the Hairpin Loop example in the lower pane of Figure 21.4. Examine the hairpin loops the oligo contains. Click the Next button (>>) to view the next hairpin loop. Click the Previous button (<<) to view the previous hairpin loop.

**Note:** To avoid any of the observed results in your final product, modify the oligo or the parameter settings and re-analyze the oligo.
Oligo Duplexes Dialog Box

This dialog box is opened by pressing the **Duplexes** button in the Oligo List dialog box or by selecting **Analyses > Oligo Analyses > Oligo Duplexes**. The Oligo Duplexes dialog box (Figure 21.5) allows you to generate all possible duplexes of selected oligonucleotides.

![Figure 21.5 Oligo Duplexes dialog box](image)

The upper pane shows oligonucleotides that are currently in the workspace. Select one oligo to analyze self-duplexes. Select any two oligos to analyze their cross-duplexes. Options in the dialog box are summarized as follows:

- **Analyze**: analyzes duplexes of the selected oligos when one or two oligos are selected, click to generate. *At least one, but no more than two oligos, must be selected to enable the Analyze button.*
- **Save Results**: Save results to the database.
- **Add New**: opens the Oligo Editor to enter a new oligo for the Oligo Duplexes workspace.
- **Remove**: removes selected oligos from the workspace.
- **Database**: adds oligos to the workspace from the oligo database.
- **Oligo List**: adds oligos to the workspace from the Oligo List.
- **dG Temperature**: enter the temperature in degrees Celsius to be used for calculating free energy values.
- **Stem Length**: enter the minimum acceptable number of base pairs in a hairpin or dimer stem.

The display pane at the bottom shows the dimers found in oligos selected in the top display window. The stem length and dimer dG are shown. Use the scroll bar to move left or right along the dimer. Click the **Next** button (>>) to view the next dimer. Click the **Previous** button (<<) to view the previous dimer.

**Note:** In the graphical depiction of duplexes, vertical lines indicate the primary interaction, based on the stem length set and plus symbols indicate secondary interactions. The greater the dG value, the weaker the interaction; secondary interactions are not considered in the dG calculation.
ORFs and Motifs Search

You can use Vector NTI tools to search for ORFs and user-defined motifs on DNA molecules. Both searches are launched from setup dialog boxes where you define parameters for the search. To initiate either search, in a Molecule Display window select **Analyses > ORF** or **Analyses > Motifs** or click the **Display Setup** button ( ) and select **Display Setup** from the drop-down menu. In the Display Setup dialog box, check the Motifs or ORFs checkbox to activate the **Motifs Setup** or **ORFs Setup** button, opening the corresponding dialog box. The parameters in these boxes are discussed in depth in Chapter 18.

After setting parameters, click **OK** to launch either search.

ORFs Search Results

ORFs search results are displayed in a Text Pane folder, listing the ORF regions meeting the criteria specified in the search. ORFs are displayed as black arrows in the Graphics Pane (Figure 21.6). Phase indicators in the left margin of the Sequence Pane show the phases of the ORFs. For the direct strand, the phase is determined by the position of the first select nucleotide relative to the beginning of the molecule. For the complementary strand, the phase is determined by the position of the last selected nucleotide relative to the end of the molecule. To eliminate the ORF arrow display, return to the Display Setup dialog box and uncheck **ORFs Setup**.

![Figure 21.6 ORFs search results](image)

Ultimate ORF Blast

Invitrogen has generated a database of proprietary ORFs, the Invitrogen Ultimate™ ORF collection. With the Ultimate ORF BLAST web tool in Vector NTI Advance, you can BLAST your fragments obtained from an ORF search against the Ultimate™ ORF collection. Using this tool, you may determine that your sequence has been previously characterized and is commercially available from Invitrogen.

To launch a search using Ultimate ORF Blast, select the target sequence by left-clicking and dragging the mouse across nucleotides in the Sequence Pane, or locating the folder in the Text Pane (under the **Open Reading Frames** subfolders) containing the target sequence. Click on the target sequence. With the target sequence selected, select **Tools > Ultimate ORF Blast**. The Sequence Data dialog box opens, allowing you to set the search parameters:
Search and Analysis Tools

Chapter 21

- **Range** selects the range of the target sequence: the entire sequence, a selected sequence or a selected CDS.
- **Strand** selects either the direct or complementary strand.

Click **OK** to begin the search or **Cancel** to abort the search process.

The Ultimate ORF Blast tool copies your target sequence and pastes it into the Ultimate™ ORF Browser Sequence Text entry box located on the Invitrogen webpage, which opens automatically when you launch this tool.

Select the animal species to search the sequence query against, checking the **Human** and/or **Mouse** checkbox located above the Sequence Text entry box. To proceed with your search, click the **BLAST** button. The Blast dialog box appears, informing you that it is blasting your sequence.

Once your search is completed, the Ultimate™ ORF Browser displays your search results. If your search identifies similar ORFs, the target sequence, or Ultimate™ ORF clone, is commercially available through the Invitrogen website, [http://www.invitrogen.com](http://www.invitrogen.com).

**Motifs Search Results**

Motifs found in a motifs search are displayed in a Motifs folder in the Text Pane and concurrently on the molecule map in the Graphics Pane (not shown). The shortcut menu that opens from a selected motif lists three commands specific to motifs:

- **Motif Site Properties** displays the site and the percent similarity of the query motif with that found on the molecule at that site
- **Add Motif Site to F(eature)Map** opens the Molecule Feature dialog box where you can name and categorize the motif before adding it to the molecule.
- **Find Motif site** (in the Graphics Pane) positions the caret at the motif site on the molecule map.

**Mutagenesis**

In a DNA/RNA Molecule Display window, you can search for “silent mutations” which, in any selected region, do not affect amino acid translation of the selected DNA but result in presence or disappearance of one or more restriction sites. You can search separately for direct or complementary sequence strand mutations.

To search for the silent mutations make sure that Display Setup contains Restriction Map options for an at least one restriction enzyme. Select a region of a molecule in the display window and choose **Analyses > Mutagenesis > Direct (or Complementary) Strand**.

In the Choose Restriction Enzymes dialog box that opens, select one or more restriction enzyme sites to be analyzed. Select at least one enzyme.

**Note:** You can only analyze restriction sites for the enzymes listed in the dialog box. To modify the enzyme selection, select **Analyses > Restriction Analyses > Restriction Sites** to open the Restriction Map Setup dialog box where you can modify the list of RENs in use. For a discussion of Restriction Map Display Setup, see Chapter 18.

Click **OK** to initiate the mutagenesis search. Vector NTI analyzes the selected region of the molecule and attempts to generate suitable silent mutations. The reading frame for amino acids is defined by the start of the selected region so that the first three nucleotides of the selected region form the first codon.

**Note:** The program is able to find both “single” (just one nucleotide altered) and “multiple” (several neighbor nucleotides altered) mutations for any elementary event (appearing and/or disappear-
ing of at least one site) significantly widening the set of possible solutions compared to just "single"-mutation analysis.

The analysis results are listed in the Mutagenesis folder (Figure 21.7) of the Text Pane of the display window.

![Figure 21.7 Mutagenesis folder](image)

The folder contains a list of mutation options that result in the appearing and/or disappearing of at least one restriction site. The options are sorted by the position of the first altered nucleotide. If you selected the complementary strand option, mutation coordinates on both complementary and direct strands are listed.

To identify a mutation region on the molecule for each mutation option, set the cursor on a numbered line under the Region line. Click the Find button ( ), the CTRL + F keys, select Edit > Find Mutation, or the corresponding command on the shortcut menu. This locates the nucleotides to be altered according to the mutation option.

**Translation**

Vector NTI offers a number of translation options, both for translating in the Sequence Pane of the Molecule Viewing window and for translating into a new molecule.
Translating in the Sequence Pane

A variety of Sequence Pane translation options are available under the Analyses drop down menu by choosing **Analyses > Translation > In Sequence Pane > <translation option>** (Figure 21.8):

- **Direct Strand** – displays a translation of the direct strand selection in the current frame
- **Complementary Strand** – displays a translation of the complementary strand selection in the current frame
- **6 Frame Translation** – displays six translations of the selection, three direct strand and three complementary strand translations, in the Sequence Pane
- **Feature(s)** – displays the translation of the selected Feature(s)
- **ORFs** – displays translations of ORFs displayed in the Graphics Pane according to filtering parameters that you designate

**Notes:**
- Translations of the direct strand appear above the sequence; translations of the complementary strand appear below the sequence.
- You can select three- or one-letter amino acid codes through the Display Setup dialog box.

To clear Sequence Pane translations, do one of the following:

- Click the **Erase Translations** button ( )
- Choose **View > Clear All Translations** from the drop-down menu or the corresponding command from the shortcut menu.

Translating into a New Protein Molecule

You can translate a DNA/RNA molecule selection into a new Protein molecule stored in the Vector NTI Database. The selected portion of a DNA/RNA molecule can be translated (direct or complementary strand) and saved in the Vector NTI Database by choosing the **File > Create New Sequence > Using Translation Procedure (Protein) > <strand option>** from the drop-down menu in the Vector NTI Molecule Viewing window.
Other options for translating DNA/RNA selections into new protein molecules are available under the Analyses drop-down menu by choosing **Analyses > Translation > Into New Protein** <trans> <translation option> (Figure 21.9):

- **Direct Strand** – creates a new protein molecule from the direct strand translation of the current selection
- **Complementary Strand** – creates a new protein molecule from the complementary strand translation of the current selection
- **6 Frame Translation** – creates six new protein molecules, one for each of the three direct and three complementary frame translations
- **Feature** – creates a new protein molecule from the translation of the selected feature
- **CDS with Splicing** – creates a new protein molecule from an Intron/Exon CDS feature by joining the Exons and then translating

**Note:** Protein molecules created by translation display a table of codon usage in the Analysis folder in the Text Pane of the Molecule Viewing window.

**Setting the Genetic Code for DNA/RNA Translations**

In Vector NTI, Standard genetic code is used for translations by default. However, you can choose from among a list of alternate genetic codes or create your own custom genetic code to use for translations.

To change the genetic code used in translations, select **Analyses > Translation > Set Genetic Code** from the drop-down menu in the Vector NTI workspace or Molecule Viewing windows. The Genetic Codes dialog box appears (Figure 21.10):

**Figure 21.10 Genetic Codes dialog box**

From the drop-down list at the top of the Genetic Codes dialog box, select the type of genetic code you want to use for translations (Figure 21.11):

**Figure 21.11 Genetic Codes dialog box code selection drop-down list**

If the desired genetic code type is not available on the drop-down list or if you want to create your own custom genetic code, click the **New** button on the Genetic Codes dialog box (Figure 21.12):

**Figure 21.12 Naming the new genetic code**
In the New Genetic Code dialog box, type in the name for the new code you are about to define and click the OK button to continue. The Genetic Code dialog box now displays the name of your new code and the buttons for defining the new code are enabled (Figure 21.13):

To define your custom genetic code, perform the following steps:

- Highlight the amino acid you want to change in the table on the left side of the Genetic Codes dialog box
- Press the appropriate amino acid button on the right side of the dialog box to redefine the highlighted box in the table
- Repeat the above two steps for each amino acid you want to change in the table
- Click the OK button to save your custom genetic code and designate it as that to be used for translations

When a custom genetic code is selected in the Genetic Codes dialog box, the Delete button is enabled, allowing you to delete the custom created code from the drop-down selection list.

**Note:** When a non-standard genetic code is displayed in the Genetic Codes dialog box, amino acids that differ from the Standard code appear shaded, as is seen in the following figure (Figure 21.14):
Back Translation

The Back Translation feature of Vector NTI Advance allows you to obtain a DNA sequence from a protein sequence by reversing the translation process. You can select from a list of codon usage table options or configure your own custom codon usage table. You can also set the display to show degrees of ambiguity in the Back Translation, from most ambiguous to most probable, as defined in the selected translation table.

Opening the Back Translation Dialog Box

You can launch the Back Translation dialog box (Figure 21.15) either from a Protein Molecule display window or from the Vector NTI Suite program files on your hard drive or Start menu as follows:

- From a Protein Molecule Display window, choose **Analyses > Back Translation**.
- From the Start menu, choose **Programs > Invitrogen > Vector NTI Suite 10 > Utilities > Back Translation**.
- From the Vector NTI Suite program files on your hard drive, double-click on the Back Translation executable file (BT.exe).

If you open the Back Translation dialog box from a Protein Molecule display window, the dialog box opens with that protein displayed. If you open the Back Translation dialog box from the Start menu or Vector NTI program files, an empty Back Translation dialog box opens. To populate the empty Back Translation dialog box, select **File > Open** from the menu to open previously saved protein documents or drag and drop a Protein Molecule from the Vector NTI Protein Database into the Back Translation dialog box.

![Figure 21.15 Back Translation dialog box](image)

The protein sequence appears in the upper pane of the Back Translation dialog box in a format similar to that in the Sequence Pane of a Protein Molecule Display window. The back translation appears in the lower pane of the Back Translation dialog box. You can change the Codon Usage and Translation tables used for the back translation, which you can display in one or three-letter code and to various degrees of degeneracy.
Codon Usage Tables

The Back Translation feature in Vector NTI allows you to specify the codon usage table you want to use for the back translation. You have the option of choosing the codon usage table from a list of many thousands of stock codon usage tables or of creating and using your own custom codon usage table.

Back Translation also has the ability to detect and repair damaged/corrupt codon usage tables.

Changing the Codon Usage Table

The current Codon Usage Table used for back translation appears in the Codon Usage Table section below the sequence. To change the current codon usage table, click the Change button or choose Table > Change Codon Usage Table from the menu to open the Codon Usage Tables dialog box (Figure 21.16):

- **Current Table Name** – shown in the grey area above the codon usage table values. A list of codon usage table choices appears on the left side of the dialog box. Highlight the table of interest to select it. The selected codon usage table data appears on the right.

**Important:** A newly selected table must be registered as the new current table. See the Use Table bullet below.

You can navigate through the codon usage table list by highlighting an entry in the list and typing the first letter (or first few letters, quickly), using the keyboard, of the name of the codon usage table you want to move to. The list scrolls to that area.

- **Filter** – provides another method of selecting a codon usage table. Type the name or part of the name of the codon usage table in the Filter text box and click the **< Apply Filter** button to select it. Codon usage table names containing the entered text display in the list; any table names not containing the entered phrase are filtered out of the list.

- **Create Table** – clears the codon usage data from the right side of the dialog box. To create a new, custom codon usage table, enter the desired codon usage table, enter the desired codon usage data, designate a new name (or the name of an existing codon usage table if you want to overwrite that table) in the **Table Name** field. Click the **Save Table** button to save the new table to the list.
Note: You can also “create” a new table by modifying an existing table. To modify an existing table, display the table and replace the appropriate displayed codon usage values.

- **Copy + Rename** – allows you to copy the selected table and designate a new name for the copy.
- **Paste Table** – allows you to paste in a new codon usage table from the Clipboard. The table must be pasted in the format found at the following Codon Usage Database site: www.kazusa.or.jp/codon. At this site, navigate to the table of interest, highlight the table, beginning with the table name and ending with the last codon/codon information listed in the table (you can also highlight only a portion of the codon information, if you do not want the entire table). In the Vector NTI Codon Usage Table, click the Paste button to paste in the table. You are asked to confirm or designate a name, depending on whether or not you included the table name as part of the copied text.
- **Save Table** – allows you to save to the list a newly created or modified codon usage table.
- **Delete Table** – deletes the selected table from the list.
- **Use Table** – registers the table as that to be used in the Back Translation dialog box. When you press this button, the Codon Usage Tables dialog box closes and the selected table appears in the Codon Usage Table field of the Back Translation dialog box.
- **Important**: If you do not click the Use Table button to register a newly selected table, when you close the Codon Usage Table dialog box, the table choice in the Back Translation dialog box defaults to the previously registered codon usage table.

**Repairing Corrupt Codon Usage Tables**

When you open the Back Translation dialog box, an automatic scan of the codon usage tables is performed. If the tables are undamaged and no corruption is detected, the back translation dialog box opens normally, as described above. If Back Translation detects corruption in the codon usage tables file, a message dialog box appears offering you the option of repairing the codon usage tables immediately or repairing them at a later time. If you choose Yes, Back Translation repairs the tables immediately. If you choose No, you can execute the repair command manually at any time by choosing Table > Repair Corrupted Tables from the Back Translation menu.

**Translation Tables and Codon Ambiguity**

In the Back Translation dialog box, you can vary the level of codon ambiguity by sliding the slider from Least Degenerate to Most Degenerate or any setting in between. Ambiguity is displayed to the degree set by the slide bar as determined by the designation translation table.

By default, the Standard genetic code translation table is selected in the Back Translation dialog box. To change the Translation table, select from among the options in the drop-down box. For more information on genetic codes, see http://www.ncbi.nih.gov/Taxonomy/Utils/wprintgc.cgi?mode=c.

**Back Translation Copy Options**

The resulting back translation sequence can be copied to the Clipboard by clicking the Copy All or Copy Selection buttons at the bottom of the screen. The resulting sequence can be pasted in the Vector NTI Sequence Editor (see Chapter 20) to create a new molecule in Vector NTI. If you copy using the Camera tool, you have the choice of copying the original sequence or the
translation to either the Clipboard or to an external text file. Choose Camera on the Main Menu to launch the Camera dialog box (Figure 21.17):

Figure 21.17 Back Translation Camera dialog box

- **Pane** – Select the pane you want to copy.
- **Range** – Select the part of the sequence you want to copy, all or selection, if you have made a selection.
- **Copy To** – Select the location for the copy.

**Restriction Analyses**

**Restriction Fragments/Restriction Sites**

From an open DNA/RNA Molecule Display window, you can calculate restriction fragments that are the result of cutting the molecule with any subset of enzymes constituting the restriction map.

With a Molecule Display window open, select **Analyses > Restriction Analyses > Restriction Fragments**. In the dialog box that opens, all enzymes are selected by default. Click **Unselect All** and select at least one restriction enzyme in the list for “digesting” your molecule.

**Note:** To modify the enzyme you can use for calculating restriction fragments, select **Analyses > Restriction Analyses > Restriction Sites** to open the Restriction Map Setup dialog box. Here you can change the list of RENs in use. For a discussion of Restriction Map Display Setup, see Chapter 19.

Vector NTI calculates the appropriate restriction fragments and lists the results in the Restriction Fragments folder of the Text Pane.

To easily identify a restriction fragment on the molecule map in the Graphics Pane, set the cursor on the appropriate line in the Restriction Fragments folder. Click the **Find** button ( ), or select **Edit > Find Fragment** or the corresponding command on the shortcut menu.

**RFLP Analysis**

RFLP (Restriction Fragment Length Polymorphism) analysis detects variations in the banding patterns of different DNA fragments that have been subjected to restriction enzyme digestion. These restriction fragment length differences originate from polymorphic regions on the DNA and can be used as markers on both physical and genetic linkage maps. Such differences in RFLP profiles have revolutionized criminal investigations and have become powerful tools in the identification of individuals in paternity and maternity cases, population genetics and in the diagnosis of a variety of diseases.
**RFLP Analysis Dialog Box**

Vector NTI's RFLP Analysis function performs a virtual RFLP assessment for two selected molecules and any number of selected restriction enzymes. To begin an RFLP analysis, choose Analyses > Restriction Analyses > RFLP from the drop-down menu in a Molecule Viewing window. The Select Molecules and Enzymes for RFLP Analysis dialog box appears (Figure 21.18):

- In the Source Molecules section, select a Database subset and highlight the two molecules you want to analyze. The highlighted molecules move to the Selected for RFLP section on the right side of the dialog box. To remove a molecule from the Selected for RFLP section, click on it to un-highlight it in the Source Molecules section.

- In the Source Enzymes section, select an Enzyme subset and highlight the enzyme you want to use for the analysis. The selected enzyme moves to the Selected for RFLP section. To remove an enzyme from the Selected for RFLP section, click on it to un-highlight it in the Source Enzymes section.

**Note:** Only one enzyme may be selected for this analysis.

- Click the Calculate button to initiate RFLP analysis.

**RFLP Analysis Results**

Unfiltered results of the RFLP analysis for the selected molecules and enzymes are displayed in the RFLP dialog box (Figure 21.19):
The RFLP results dialog box shows the resulting digestion products for the two analyzed molecules with each of the selected restriction enzymes.

- **Use Filter** – Initially the results are unfiltered (i.e. results for all analyzed restriction enzymes are shown). You can subsequently filter the results by checking the Use Filter box, specifying the filtering parameters in the Number Of Sites In sections and clicking the Apply button (Figure 21.20):

![Figure 21.20 Filtered RFLP results](image)

- **Results** – Double-click on the enzyme entry in the RFLP Results section to view the digestion fragments for the analyzed molecules with that enzyme (Figure 21.21).

![Figure 21.21 Digestion fragments dialog box](image)

- **Create Gel** – Launches the Gel Setup dialog box so you can perform a gel analysis on the RFLP results for an enzyme analysis. For more information on Vector NTI’s Gel Analysis function, see Chapter 25.

### Finding Common Non-Cutter Restriction Enzymes

The Find Common Non-Cutter Enzymes tool in Vector NTI performs a restriction analysis on two or more selected DNA/RNA molecules for one or more selected restriction enzymes. The tool reports non-cutting enzymes from the selected list that are common to all molecules selected for analysis.

To launch the Non-Cutting Enzymes tool, choose Analyses > Restriction Analyses > Find Common Non-cutter RENs from the drop-down menu in a Vector NTI DNA/RNA Molecule viewing window. The Non-Cutting Restriction Enzymes dialog box appears with the active molecule selected for analysis (Figure 21.22):
**Search and Analysis Tools Chapter 21**

Figure 21.22 Non-cutting Restriction Enzymes dialog box

- **Molecules** – Lists the molecules in the Vector NTI Database DNA/RNA Molecule subset selected in the **Subset** drop-down box.
- **Enzymes** – Lists the enzymes in the Vector NTI Database Enzyme subset selected in the **Subset** drop-down box.
- **Selected for Analyses** – Lists the DNA/RNA Molecules and Enzymes you have selected for Non-cutter Enzyme analysis. Allows you to set an upper limit for cut sites; enzymes that cut more than the designated number of times are ignored.

**Note:** Initially, the **Selected for Analyses** section of the dialog box is not populated.

Highlight the molecules you want to analyze in the **Molecules** section and click the > button to enter those molecules into the **Selected for Analyses** section. Click the >> button to enter all the molecules in the selected subset into the **Selected for Analyses** section.

Highlight the enzymes you want to use for the analysis in the **Enzymes** section and click the > button to enter those enzymes into the **Selected for Analyses** section. Click the >> button to enter all the enzymes in the selected subset into the **Selected for Analyses** section.

Click the **Start** button to begin the analysis. When the analysis process completes, the Preview: Non-Cutting Enzymes Report window displays (Figure 21.23):

Figure 21.23 Non-Cutting Restriction Enzymes Report preview window
Non-cutting enzymes common to the molecules analyzed are listed in the Preview: Non-Cutting Enzymes Report window. From the Preview: Non-Cutting Enzymes Report window, you can print the file directly, copy it with the Camera feature to the Clipboard or file, or save it directly to an external file.

Restriction Report

You can create a restriction analysis report for the active molecule with all the enzymes in your Vector NTI Enzyme Database using Vector NTI's Restriction Report analysis tool. To create a restriction analysis report for a molecule, open the molecule in a Molecule Viewing window in Vector NTI and choose Analyses > Restriction Analyses > Restriction Report from the Main Menu. When the analysis is completed, a summary of the results appears in the Report Preview window (Figure 21.24):

![Restriction Report](image)

*Figure 21.24 Restriction Report results for M13mp11*

The Restriction Report result has two sections, Cutters and Non-Cutters. The Cutters folder shows, for each cutting enzyme, the enzyme name, recognition sequence and cut site position(s) on the analyzed molecule. The Non-Cutters section shows the list of the enzymes that do not cut the analyzed molecule.
Introduction

This chapter describes the essentials of creating new molecules with Vector NTI. Features of two important dialog boxes used for molecule creation—the Construct/Design Molecule dialog box, and the Molecule Editor—are discussed.

This chapter covers the following points:

- descriptions of molecule types: basic and constructed
- differentiating in Vector NTI between molecule Construction and Design
- preparing fragments for Construction using Fragment Wizard
- using the Construct/Design Molecule dialog box as a tool for molecule creation
- modifying component fragments termini with biochemical operations
- using Molecule Editor to prepare molecule fragments
- importing molecules

Molecule Types: Basic And Constructed

Molecules in Vector NTI can be of two fundamental types: basic and constructed.

**Basic** molecules are not built from component fragments. Their sequences, Feature map, and restriction maps of unsequenced regions (DNA molecules only) are entered by the user or imported from other databases.

**Constructed** molecules are built from one or more fragments. For DNA molecules, these can be fragments of existing DNA molecules, linkers, adapters, etc. The Feature map and nucleotide sequences of parent molecules are automatically transferred to a constructed DNA molecule. You can add features to the Feature map in addition to those obtained from the ancestors. Constructed DNA molecules are created by Construction or Design, Vector NTI terms for recombinant processes that are differentiated below.
Protein molecules can be constructed by translating a certain region of a DNA molecule. A constructed protein molecule does not receive any Feature map from its parent DNA molecule.

**Methods of creating new DNA/RNA molecules**

In Vector NTI, there are several different ways of creating new DNA/RNA molecules:

**Basic Molecules:**
- **Importing molecules or sequences** (including their feature tables) in the following formats: GenBank, GenPept, EMBL, SWISS-PROT, FASTA, and ASCII text. You can also import molecule data from "quasi-GenBank" files (that is, GCG's "GenBank" output files). Molecule import is discussed on page 182.
- **Creating new molecules “from scratch.”** You can define a molecule’s nucleotide or amino acid sequence by hand or paste it from the clipboard and enter the sequence as a new molecule, describing the feature and restriction maps if you wish. This is described in Chapter 19 (see page 237).

**Constructed Molecules:**
- **Creating new DNA/RNA or Protein molecules by splicing** together the exons of an Intron-Exon join feature of another molecule. In the case of Protein molecule creation, the spliced molecule is subsequently translated and stored in the VNTI Protein Database. A molecule created by this method is considered a constructed molecule; it is a child molecule of the molecule it was spliced (or spliced and translated) from.
- **Construction of new DNA/RNA molecules** from compatible components fragments from other molecules, linkers, adapters, etc. In Construction, the fragments and restriction sites are entirely defined by the user.
- **Design of new DNA/RNA molecules** from components also in a user-defined fragment list. In Design, however, instead of personally choosing the restriction sites, methods of terminus modification, etc., the user turns the Design process over to Vector NTI. The software takes advantage of the best possible restriction sites and genetic engineering techniques to design the recombination process.
- **Translation of new Protein molecules** from selected portions of existing DNA/RNA molecules. A protein molecule created by this method is considered a constructed molecule; it is a child of the DNA/RNA molecule from which it was translated. For more information on creating new protein molecules using various translation procedures, see Translating into a New Protein Molecule, page 293.

**Note:** For creating new molecules, Vector NTI must be in Molecule Editing mode, the default mode when the Molecule Display window opens.
Creating New DNA/RNA or Protein Molecules Using Splicing

For molecules with a marked intron/exon CDS feature (denoted by the join statement in the molecules GenBank file, see Figure 22.1), you can create a new molecule from the intron/exon feature by splicing out the introns and joining the exons.

To create a new molecule by splicing the intron/exon feature of another molecule, open the molecule with the intron/exon feature in a Molecule Viewing window, select the intron/exon feature in the Graphics Pane and choose File > Create New Sequence> Using Splicing (DNA/RNA/Protein) from the drop-down menu. The Create New Molecule(s) by Splicing dialog box appears (Figure 22.2):

In the Source section of the Create New Molecule(s) by Splicing dialog box the following information is defined:

- **Molecule Name** – this is preset according to the name of the opened molecule
- **Feature Position** – predefined according to the molecule selection when the Create New Molecule(s) by Splicing dialog box is launched
• **Feature Type(s) To Splice** – initially, the Splice Feature Types box is empty; you must select the feature type(s) to splice from the Add Features dialog box. To add a feature type, click the plus sign button to the right of the Splice Feature Types box (Figure 22.2). This opens the Add Features dialog box (Figure 22.3):

![Add Features dialog box](image)

**Figure 22.3 Add Features dialog box**

• **Strand** – direct or complementary

The Target section indicates the type for the resulting spliced molecule. If Protein is selected, the feature is spliced and then translated into a Protein molecule.

In the Results Should Be From section, you can define the number of spliced products you want generated:

- **Min # of Exons** – specify the minimum number of exons to splice to create a product.
- **Max # of Exons** – specify the maximum number of exons to splice to create a product.

Vector NTI creates the range of all possible spliced products between the specified minimum and maximum number of exons. If you want only the full length product (all exons spliced), enter the total number of exons in both the Min # of Exons and Max # of Exons fields. Unchecking the Results Should Be From box results in the generation of all possible products from one exon in length to the maximum possible number of joined exons for the selected molecule feature.

### Tools for Creating New Constructed and Designed DNA/RNA Molecules

Vector NTI has three important tools for creating new constructed molecules: the Fragment Wizard, the Construct/Design Molecule dialog box and the Molecule Editor. The process of creating a molecule generally follows these steps, in an active Graphics Pane:

1. Define the component fragments for the recombinant molecule using Fragment Wizard.
2. Add defined fragments to the Goal Molecule Definition List.
3. Open the Construct/Design Molecule dialog box where you set construction parameters.
4. Name, select data and describe the new molecule.
5. Verify and edit, where needed, the component fragments in the Goal Molecule Definition List.
6. Initiate molecule construction.

If construction is successful, your molecule is entered into the database and is opened in a new Molecule Display window.

The Molecule Editor is the tool for editing existing database molecules and for creating new basic molecules from scratch. For details, refer to page 237.
Construction Tips:

- Before you begin, open display windows for all of the molecules that are fragment sources. If you have several display windows open concurrently, the source molecule for your fragment must be the topmost window.
- It is sometimes convenient to work with two Molecule Display windows on display at the same time. Select Window > Tile Vertical. For each window, resize the Graphics Pane and molecule maps using the split bars, scroll bars and Zoom buttons to optimize viewing of the graphics. Click in a display window to activate it; an active window is designated by a dark blue Title Bar.

Describing Component Fragments in the Fragment Wizard

The Fragment Wizard guides you through several steps required to fully define molecule fragments suitable for construction or design. While in view, it allows you to continue to work in the Graphics Pane to select the fragment, while providing visual feedback and explanation of the actions.

To define a fragment using Fragment Wizard, open a display window for the molecule to be the source of a fragment.

Activate the Graphics Pane and click the Add Fragment To Goal List button ( ), or select List > Add Fragment to Goal List, opening the Fragment Wizard dialog box (Figure 22.4):

![Fragment Wizard dialog box](image)

Figure 22.4 Fragment Wizard dialog box

The Fragment Wizard screen changes to display the entry fields and descriptions corresponding to the current step in the fragment definition process.

At any point in the process, you can choose to Cancel, go Back a screen, go to the Next screen, or Finish the process using the appropriate buttons. (Finish goes directly to the last step, skipping all the remaining steps, if any. Fragment Wizard goes through all the steps you skipped and supplies the missing information using default values and the coordinates of the current selection in the Display window.)

In the first screen (Figure 22.4):

Select the fragment type and process: Construction fragment, Design Recipient fragment and Design Donor fragment. Since this chapter covers Construction, select Construction Fragment.

Design donor and design recipient fragments are discussed in Chapter 24.

When Construction fragment is selected, you can specify that the fragment should be inverted in the constructed molecule by checking the Insert Inverted box. Click the Next button to proceed.
In the second screen (Figure 22.5):

Specify the 5’ terminus of a new fragment. The options are:

- Set to a position on the molecule map (in one of the two following ways)
  1. Enter the coordinate in the text box or CLICK in the Graphics Pane of the display window. In the latter case, the option button will be selected automatically and the caret position is displayed in the position field of the dialog box and on the Status Bar. Use left or right arrows to reposition the caret. **Tip:** If, when you click in the Graphics Pane, the entire molecule is selected (as indicated by the wire-frame), click in the Sequence Pane to deselect, and begin again.
  2. Specify the start of the molecule
     To set the 5’ terminus at the start of the molecule, select the Start of Molecule option in the dialog.

- Specify a restriction site.
  Click on the site or its label in the Graphics Pane. The name of the restriction endonuclease and the position of the site are displayed in the Set to a Restriction Site box.

**Note:** If the Fragment Wizard screen limits your view of the molecule map, click on its title bar and drag it out of the way.

![Figure 22.5 Specifying the 5' terminus of a new fragment](image)

Click the **Next** button to proceed.

In the third screen (Figure 22.6):

Specify the 3’ terminus of the fragment. Options are:

- Specify a position on the molecule map (in one of the two following ways)
  1. Enter the position in the text box or SHIFT + CLICK in the Graphics Pane on a position, restriction site or label, automatically selecting this option. The 5’-3’ selection is now indicated with a wire-frame in the Graphics Pane and in the set selection box on the Status Bar. Reposition the 3’ end using SHIFT + the left or right arrows.
  2. Specify the Start or End of the Molecule
     The selection is indicated with a wire-frame and on the Status Bar.

- Specify a restriction site.
  SHIFT + CLICK on the site or its label in the Graphics Pane.
Tip: The SHIFT key must be held down for all actions in the Graphics Pane for setting the 3’ end. If you neglect to do this, your actions will change the 5’ terminus and Fragment Wizard will revert to screen 2.

Figure 22.6 Specifying the 3’ terminus of the fragment

Click the Finish button, completing the fragment definition.

Review the description of the selected fragment in the New Fragment message box (Figure 22.7).

Figure 22.7 New Fragment message box

If you made a mistake, press Cancel to return to the Fragment Wizard. Otherwise, click the Add to List button, adding the fragment to the Goal Molecule Definition List and closing the Fragment Wizard.

Now go to the Molecule Display window for the molecule source of the second Construction fragment and define it using the Fragment Wizard in the manner just described. After reviewing your fragment as before in the New Fragment dialog box, enter it into the Goal Molecule Definition List.

The Goal Molecule Definition List (Goal List)

The Goal Molecule Definition List is a list of fragments that will be combined to create a new molecule.

You can add fragments to the Goal Molecule Definition List using the Fragment Wizard as just described or you can add them directly from the Construct dialog box.

To see the Goal Molecule Definition List, click the Open Goal List button on the Main Toolbar or choose List > Molecule Goal List. At the bottom of the Lists dialog box (Figure 22.8)
that opens, the Fragments display lists the fragments currently on the Goal Molecule Definition List.

Figure 22.8 Lists dialog box

Constructing a New Molecule

The Construct & Design Molecule Dialog Boxes

In these dialog boxes, you can describe the new molecule to be constructed (or designed) and saved into the Vector NTI database. Also you can open it for an existing constructed or designed database molecule to reconstruct/redesign it, that is, after changing its component fragments.

To open the Construct or Design dialog box, first open the Lists dialog box from a Molecule Display window by clicking the Open Goal List button ( ) or selecting File > Create New Sequence > Using Construct/Design Procedure (DNA/RNA) or selecting List > Molecule Goal List. Click the Run button on the Lists dialog box to display the Construct or Design Molecule dialog box. The dialog box is automatically loaded with the new molecule described through Goal Molecule Definition List.

Note: There are two operational modes—a mode for molecule Construction and mode for molecule Design. The appropriate mode is set automatically when you designate the component fragment type in Fragment Wizard. The title of the box “Construct Molecule” or “Design Molecule” reflects the mode. (If this dialog box is opened before the method is designated, the box title is also Construct Molecule.) Because Construction is the subject of this chapter, we will refer to it as Construct Molecule dialog box (Figure 22.9).

Figure 22.9 Construct Molecule dialog box

Enter or select the dialog box fields as follows:

- **Name** - enter a new name for the molecule to be constructed
- **Circular/Linear** - the form of the molecule.
- **DNA/RNA** - molecule type
- Click the General Info button to enter molecule description information:
The Component Fragments Field

The Fragments field in the Lists dialog box contains the Goal Molecule Definition List. Constructed molecules consist of one or more component fragments that can be verified, entered, deleted, repositioned or edited in this field. In addition, you can add new fragments of five types, described below in the section on adding new fragments.

To perform an operation on a fragment, select it and press one of the following enabled buttons:

- **Edit**: This action can also be initiated by double-clicking the fragment. In the dialog box that opens, you can edit fragment information. The specific dialog box appears depends on the type of the fragment you have highlighted. Five fragment types (described below in the section on adding new fragments) each have their own Edit dialog box.
- **Delete**: The selected fragment disappears from the list.
- **Reposition a fragment**: To change the position of a fragment in the list, highlight the fragment you want to move, and click the **Up** or **Down** buttons to change its position in the list.

The molecule will be constructed from the fragments in the order shown, so changing the positions in the list changes the resulting molecule. The fragment listed first is always considered the “recipient” fragment.

Adding New Component Fragments

To add a new fragment to the list of component fragments, click the **Add** button above the Fragments field and choose the desired fragment type from the drop down list. In the corresponding Fragment Editor dialog box that opens, you can enter information about the new fragment:

- **Dummy**: A Dummy fragment can be created for a recombinant molecule as a last resort when a DNA fragment length is known but the exact sequence is not. Residues in Dummy fragments are added as “nucleotides” only. *They will appear in a sequence as “N”.*

When you add or edit a Dummy fragment, the Dummy Fragment dialog box (Figure 22.10) appears where you can enter the number of unsequenced nucleotides the fragment contains:

![Dummy Fragment dialog box](image-url)

*Figure 22.10 Dummy Fragment dialog box*
Dummy fragments have blunt termini. A Dummy fragment must be inserted into a vector with compatible linkers or adapters flanking the sequence. To add REN sites to ends of a Dummy fragment, see the Linker section below.

- **Sequence**: Sequence fragments are user-specified oligonucleotides. When you add or edit a Sequence fragment, the Sequence Fragment dialog box appears (Figure 22.11):

  ![Sequence Fragment dialog box](image)

  In the Nucleotide Sequence field, enter an oligonucleotide up to 120 units long. With the radio buttons, indicate direct or complementary orientation.

  Sequence fragments have blunt termini. A Sequence fragment must be inserted into a vector with compatible linkers flanking the sequence. To add REN sites to ends of a Sequence fragment, see the Linker section below.

- **Linker**: Linker fragments are single restriction sites that can have short flanking sequences. They are used to link together longer fragments in Construction or Design. When you add or edit a Linker fragment, the Linker Fragment dialog box (Figure 22.12) appears:

  ![Linker Fragment dialog box](image)

  In the Restriction Site field, enter the name of the restriction endonuclease being used or click the **Browse** button ( ) to select from a list of RENs in the database.

  In the **Left Nucleotides** and **Right Nucleotides** fields, you can enter nucleotide sequences flanking the restriction site.

  Once the restriction site is in place, select the radio button that indicates whether the linker should be the whole restriction site or whether it will attach at the left end or the right end of another sequence or dummy fragment.

  If you click **Whole**, then both ends of the Linker are blunt. If you click **Left**, the 5’ end is blunt and the 3’ end is the restriction site. If you click **Right**, the 3’ end is blunt and the 5’ end is the restriction site.

  **Examples:**

  Left Linker *HindIII* \(^\text{A}\)TTCGA \(^\text{G}\) Right Linker *EcoRI* AATTC

  If you want restriction site Linkers to be attached on each end of a sequence to be cloned, you must generate a right and a left Linker. In the following example, the Left Linker will ligate to the 5’ end of the digested recipient fragment and the Right Linker will ligate to its 3’ end.

  ![Linker example](image)
Construction Tip: To add linkers at each end of a Dummy or Sequence fragment, they should be listed in the Component Fragment list in the following order:

- Recipient Fragment
- Linker (Left Linker)
- Dummy or Sequence
- Linker (Right Linker)

- Adapter: Adapter fragments are like linkers except that in the Adapter Fragment dialog box, you enter or edit nucleotides manually or choose the components of the Adapter Fragment from the Enzyme and Oligo Databases. Adapters with sticky or blunt ends can be designed. Adapters are often special-ordered from a synthesizer vendor, as they tend to be linkers that are “adapted” to the given experiment/need at hand.

When you specify adding an Adapter, the Adapter Fragment dialog box appears (Figure 22.13). In Vector NTI, an Adapter consists of a 5’ or 3’ left terminus, a central double-stranded part, and a 5’ or 3’ right terminus, although any part of an Adapter fragment is optional.

In the Adapter Fragment dialog box, enter these segments manually in the appropriate fields of the dialog box or click the appropriate Load button near the bottom of the Adapter Fragment dialog box to select the various Adapter parts from the Restriction Enzyme or Oligo Databases:

Once you have selected the appropriate Oligo and left and right Restriction Enzymes for the Adapter, click the Calculate Adapter button to calculate and enter the sequence into the upper portion of the Adapter Fragment dialog box (Figure 22.14):

Once you have completed your Adapter fragment design, click the OK button to add the new Adapter to the Fragment List.
Add Fragment: There are three choices available from the Add Fragment selection, adding a fragment from the molecule that is currently in focus in the VNTI workspace, adding a fragment of a molecule in the VNTI database or adding a fragment using the Fragment Editor. The first two choices will open the appropriate molecule and launch the Fragment Wizard. The third choice will launch the Fragment Editor.

Note: Although you can add molecule fragments using the Fragment Editor, it is much more convenient to add them either in advance using the Fragment Wizard—the tool specifically designed for visual selection of molecule fragments from Molecule Display windows (see page 309) or via the first two choices on the Add Fragment list, which also use the Fragment Wizard.

Fragments of existing molecules are the key part of most constructed molecules. Click the Add button and select the Add Fragment > with Fragment Editor choice, or select a component fragment in the list and click the Edit button to open the Fragment of Molecule dialog box (Figure 22.15) where you add or edit such a fragment:

![Fragment of Molecule dialog box](image)

Options in this dialog box are summarized as follows:

- Construct/Design Recipient/Design Donor: molecule fragment type
- Parent Molecule: Enter the name of the source molecule for the fragment to be added, or click the Browse button ( ) to select a molecule from the Vector NTI database.
- Inverted: Check to insert the fragment in the inverted orientation.
- Restriction Sites Counted From Nucleotide: Enter the position in the parent molecule from which restriction site occurrences should be counted. This field is important when this fragment is taken from a selection display window, where the numbering of sites in the selection display window may not correspond to the numbering of sites in the molecule as a whole. Usually the value of this field is 1.
- Termini: To describe a fragment from an existing molecule, you must describe its left and right termini. Brief descriptions of the termini are displayed next to the corresponding buttons. Click each of the <...Terminus > buttons to define the termini.
Editing the Termini of Fragment of Molecule-type Component Fragments

Click the **Left Terminus** or **Right Terminus** buttons in to open the Terminus Editor dialog box (Figure 22.16).

![Terminus Editor dialog box](image)

**Figure 22.16 Terminus Editor dialog box**

The Left (and Right) Terminus dialog box options are summarized as follows:

**Terminus Type:**
- **Restriction site:** If the terminus is a restriction site, enter its name in the text box, or click the **Browse** button ( ) to select from a list of the RENs in the database.
  
  *The restriction site chosen here determines the form of the terminus (blunt, 5’, 3’). The terminus can be further modified by biochemical operations, as described below.*

- **Nucleotide context:** If the terminus is a nucleotide context (oligonucleotide), enter the sequence in the text box.
  *This type of terminus is blunt.*

- **Position:** If the terminus is a position in the molecule’s nucleotide sequence, enter its position in the text box.
  *This type of terminus is blunt.*

- **Start / End of Molecule:** If the terminus is the start or end of the parent molecule, click the appropriate radio button.
  *This type of terminus is blunt.*

- **Not Defined (NODEF):** Click this button if the terminus is undefined.
  *A fragment with one or more NODEF termini can only be used for automatic molecule Design, discussed in the next chapter.*

**Site or Context Data**

If the terminus is a restriction site or nucleotide context, these fields are enabled where you are required to indicate where on the parent molecule the site or context can be found.

- Enter an integer value in the Number field to use that occurrence of the site or context on the parent molecule.
  *In place of a number, you can enter an area of the molecule within which the site or context occurs only once.*

- Enter a value in the Point and Region fields to look for the site or context within (Region) nucleotides of the (Point) nucleotide of the parent molecule.
Vector NTI uses the Point and Region fields only if the Number field is empty.

**Biochemical Operations**

Termini of Restriction Site type can be treated with up to three consecutive biochemical operations, performed in the sequence listed in the Terminus Editor dialog box. The operations are performed in the order in which they are listed.

These operations from the lists in the Biochemical Operations drop-down menus have the following effects:

- **None** leaves the terminus untreated.
- **Completely filled in** fills in cohesive fragment termini to make them blunt.
- **Incompletely filled in** fills in cohesive fragment termini to within a few nucleotides of being blunt.
- **S1-treated** cuts cohesive termini down to blunt.
- **Bal31 treated** cuts off a given number of nucleotides from both strands.
- **3'-5' exonuclease treated** cuts off a given number of nucleotides in the 3'-5' direction.
- **5'-3' exonuclease treated** cuts off a given number of nucleotides in the 5'-3' direction.

Click **OK** to return to the Fragment Molecule dialog box (Figure 22.15).

The For Design Only subsection of this dialog box is enabled only in Design mode and is described in Chapter 24.

Click **OK** to return to the Lists dialog box. Click the **Run** button to launch the Construct Molecule dialog box.

**The Control Fields and Commands**

In Construction mode, the following mode is enabled: **Create All Possible Constructs**...

When checked, the system creates all possible constructions (up to 24) that can be obtained from the molecule component fragments. This option allows Vector NTI to include even the parasitic constructs (inserts in the wrong orientation) in the reaction mix.

The **Open Display Windows for All** field is active only when the **Create All Possible Constructs** box is checked or the specified number of design output options is more than one. If this box is checked, and more than one molecule is created during the construction or design process, Molecule Display windows are opened for each new molecule immediately after completing the procedure. If this box remains unchecked, a new Molecule Display window opens automatically for just the molecule with the name specified in the Name field.

**Note:** In Design, display windows are not created for intermediate molecules, but only for the final option(s).

Once all parameters are set in the Construct Molecule dialog box, click the **Construct** button to initiate the process.

**Naming of Construction Recombinants**

If you start a construction task for a goal molecule with a name that causes a conflict with an existing set of construction options, Vector NTI informs you of the conflict and asks permission to delete the existing options. To avoid this, choose a different name for the new goal molecule or rename old construction options.

For additional constructs, Vector NTI creates names on the basis of the name of the "main" constructed molecule. If the name of the molecule is `<MOLNAME>`, the name of the first additional construct will be `<MOLNAME> #1`, the name of the second additional construct will be `<MOLNAME> #2`, etc.
As Construction proceeds, Vector NTI creates a special subset for the molecules, named by the program on the basis of the name of the “main” constructed molecule. If the name of the molecule is `<MOLNAME>`, then the name of the subset containing additional constructs will be `<MOLNAME> _$Construction Options.`

**Reconstructing a Molecule**

You can reconstruct a previously constructed molecule by right-clicking on the constructed molecule in the Database Explorer window and choosing **Reconstruct** from the shortcut menu. This launches the Construct Molecule dialog box with all the parameters used for the original construction of that molecule. The Reconstruct feature is useful for constructing molecules that are similar to a previously constructed molecule and for reconstructing molecules that have failed a previous construction attempt (such molecules appear with a question mark icon to the left of their names in the Database Explorer window).
Introduction to Gateway Cloning

Gateway Cloning Technology is Invitrogen’s universal cloning system that provides a rapid and highly efficient route to protein expression, functional analysis and cloning/subcloning of DNA segments. This system is based on the well-characterized bacteriophage lambda-based site-specific recombination system ($\text{att}_L \times \text{att}_R \leftrightarrow \text{att}_B \times \text{att}_P$). In overview, Gateway is a 2-step cloning process; in the first step a sequence of interest is inserted into an Entry Clone. The second step transfers the sequence into a variety of $\text{att}_B$-containing Expression Clones that can be propagated and expressed in a range of host cells for a given experiment.

TOPO Technology is a fast, efficient way to clone. The key to TOPO Cloning is the enzyme DNA topoisomerase I, whose biological role is to cleave and rejoin DNA during replication. To harness this activity, vectors are linearized and each end is conjugated with topoisomerase on the 3' phosphate. This enables fast ligation of DNA sequences with compatible ends. After 5 minutes at room temperature, enzyme is released, the ligation is complete and the recombinant molecule is ready for transformation into $E. coli$. Many Invitrogen expression vectors are adapted for one-step TOPO Cloning of PCR products in both directional and non-directional formats. Other vectors contain $\text{att}$ recombination sequences exterior to the TOPO cloning sites so that cloned inserts are ready for entry into the Gateway system.

Vector NTI Advance 10 offers an interlocking set of computational tools for planning Gateway or TOPO cloning experiments from beginning to end. These allow you to simulate recombinant strategy, as well as to validate design of your reagents before proceeding to primer ordering, bench-top operations, etc. Steps necessary to identify and screen transformants, expedite purification, fuse an insert with vector tags, etc. can be identified and modifications made before
investing time and money. Vector NTI tools can be used to plan individual experiments (single- or multi-site cloning) or to design batch cloning experiments.

<table>
<thead>
<tr>
<th>Warning !</th>
</tr>
</thead>
<tbody>
<tr>
<td>If you upgraded to Vector NTI Advance 10 from a previous version and chose to continue working with your existing database, some Invitrogen molecules needed for the Gateway Cloning feature will not exist in your database. To obtain those molecules, you must download them from the Invitrogen website at <a href="http://www.invitrogen.com">www.invitrogen.com</a>.</td>
</tr>
</tbody>
</table>

If you chose to overwrite your existing database when you upgraded, the Invitrogen molecules will be part your Vector NTI database.

**Introduction to Gateway Cloning**

This chapter describes the Vector NTI tools for *in silico* Gateway Cloning and provides you with directions for using these tools:

- **Creating Entry Clones** – To participate in a Gateway cloning reaction, a DNA sequence of interest must first be incorporated into an Entry Clone. Five ways for generating or retrieving Entry Clones are described in this chapter:

  a. **Generate an Entry Clone from an Expression Clone by BP Recombination** – This option uses a BP reaction to generate an Entry Clone employing an already existing attB-containing DNA molecule, such as a Gateway Expression Clone or a pCMVSPORT6 library. See page 323.

  \[ \text{DNA sequence of interest in an existing clone (flanked by attB sites) + pDONR vector (with attP) } \leftrightarrow \text{ Entry Clone (with attL)} \]

  b. **Create an Entry Clone by PCR Amplification and BP Recombination** – This option allows you to amplify your sequence of interest with attB-containing primers designed by Vector NTI, then perform a BP recombination using the PCR product you amplified. See page 324.

  \[ \text{PCR product (flanked by attB sites) + pDONR vector (with attP) } \leftrightarrow \text{ Entry Clone (with attL)} \]

  c. **Generate an Entry Clone by PCR and Restriction-Ligation into a pENTR Vector** – This process allows you to specify your sequence of interest and amplify it using standard Vector NTI primer design and amplification techniques. Then insert the amplicon into a pENTR vector using Vector NTI molecule construction techniques. See page 329.

  \[ \text{PCR product (no attB) + pENTR vector (with attL) } \leftrightarrow \text{ Entry Clone (with attL)} \]

  d. **Obtain a functional Entry Clone from Invitrogen** – Find Entry Clones in the Invitrogen collections. See page 331.

  e. **Generate an Entry Clone by TOPO cloning** - This option allows you to amplify your sequence of interest and clone it into a Gateway-ready TOPO vector. This is described fully in the TOPO Cloning section of this chapter. See page 323.

- **Perform an LR Recombination** – Entry Clones from any of the above methods are recombined with pDEST vectors. The resulting Expression Clones can drive expression of the sequence of interest when transformed into host cells. See page 331.

  \[ \text{Entry Clone (with DNA sequence of interest flanked by attL sites) + pDEST vector (with attR) } \leftrightarrow \text{ Expression Clone (with attB)} \]
ORF and Translation Analyses – Use Vector NTI analysis tools to verify that a DNA sequence in an Expression Clone is expressed as intended before committing bench time to subsequent steps. See page 334.

Creating a New Destination Vector – This allows you to create a user-defined destination vector, incorporating unique promoters, fusion tags, etc. See page page 335.

Selected vector + Reading Frame Cassette (with attR) → Destination Vector

Creating Entry Clones

Invitrogen offers several methods for inserting a DNA sequence of interest into an Entry Clone. Vector NTI tools can perform these recombinant techniques in silico, and additionally, Vector NTI can be used to design primers, amplify products, construct molecules, track recombinants, suggest reagents and predict the outcome.

As a general rule, any Entry Clone generated in silico should be submitted to a Gateway LR recombination in Vector NTI. (See Performing an LR Recombination on page 331.) The resulting Expression Clones should be examined by ORF analysis and translation to verify that the DNA sequence will indeed be expressed as intended.

Generate an Entry Clone in a BP reaction by Recombining an Expression Clone

A DNA sequence of interest may already be in an attB-containing DNA molecule such as an existing Gateway Expression Clone or pCMV-SPORT6 library. For information, contact Invitrogen Technical Services at the Invitrogen website http://www.invitrogen.com.

In this strategy, the sequence is extracted from the existing Gateway clone or library and recombined with a Donor Vector in a BP recombination to produce a new Entry Clone. Ultimately, the Entry Clone can be recombined into a new Expression Clone.

1. Open the Expression Clone in a Molecule Display Window, and from the menu, choose Cloning > Gateway Cloning > Create an Entry Clone > Recombine an Expression Clone by BP.

   Alternatively, in the Database Explorer, select one or more such clones and from the DNA/RNA or shortcut menu choose Gateway Cloning > Recombine an Expression Clone by BP.

2. First screen (Figure 23.1): Use the Add/Remove buttons to browse in one or more pDONR vectors and attB-containing sequences, in the upper and lower windows respectively. If the tool was launched from an open file window displaying either type of molecule, those molecules automatically appears in the appropriate pane. Click Next.

   Note: pDONR vectors for multi-site Gateway reactions contain attP sites that are different from those used in single-fragment cloning. In all reactions, the attB sequences must be compatible with
pDONR attP sites. If invalid files or incompatible vectors and sequences are entered, Vector NTI will issue an alert and return to this screen.

3. **Second screen** (Figure 23.2): An expandable preview page appears, summarizing valid Entry Clones that can be generated from the entries displayed on the first screen. Batch submissions from the first screen (e.g. a single pDONR vector and multiple compatible attB sequences) are summarized as separate Entry Clones.

4. To automatically save the results to the Vector NTI database, check the **Save results to DB** checkbox. You are offered the option to open the new clone in a Molecule Display Window. Click **Finish**.

   In molecules created by the Gateway cloning tools, a folder called **Design Description** appears in the Text Pane. It contains a brief description of the protocol and can be copied and pasted into a word document for reference during the actual experiment.

5. If the **Save results to DB** option was not checked, Entry Clones will open in Molecule Display Windows as temporary molecules that can be saved in the Database.

6. Perform an LR recombination in Vector NTI (see page 331) and examine the resulting Expression Clone by ORF analysis and translation to verify that the DNA sequence is expressed as intended (see page 334).

**Generate an Entry Clone by PCR Amplification and BP Recombination**

This strategy is a highly flexible process for entering the Gateway Cloning system by amplifying a sequence of interest using standard PCR techniques and generating an Entry Clone in a BP Recombination. This is also the approach used most often when multi-site Gateway cloning is
planned, since each fragment of the final clone must be prepared with a unique combination of attB sites. Some of the benefits to this approach to creating Entry Clones are:

- You can add additional sequences to the primers to assist in protein expression and purification.
- You can open online help topics for assistance in the placement of stop codons and maintaining reading frames.
- If reading frame errors are detected, you can return and make the appropriate edits.
- You can add primer recommendations to an Oligo Goal List from which orders can be submitted to Invitrogen's custom Oligo Synthesis service.

1. Open a Molecule Display Window for a DNA molecule containing the sequence of interest. Select the target region on the molecule and choose **Cloning > Gateway Cloning > Create an Entry Clone > Amplify Selection to Use in a BP Reaction**. (These commands are disabled if an appropriate target is not defined.) This opens the Amplification of Selected Region for BP Recombination dialog box (Figure 23.3).

This dialog box allows you to search for convenient sense and antisense primers according to your specifications and to check the uniqueness of primers. Most of the tabs and parameters in this dialog box are common to those found in the Find PCR Primers dialog box discussed in Chapter 20. However, the differences on the Primer tab are described here. For more information on setting PCR parameters in this dialog box, see page 248.

![Figure 23.3 Amplification of Selected Region for BP Recombination dialog box](image)

The options on the **Primer** tab (Figure 23.3) allow you to generate a single pair of primers that amplifies but does not extend beyond the selected sequence. It also allows you to add primer attachments for a Gateway Cloning experiment.

| Primer Tab Parameters – Amplification of Selected Region for BP Reaction Dialog Box |
|---------------------------------|--------------------------------------------------|
| Region of Analysis             | Restricted to exactly the region being amplified. Therefore, Before = After = 0 bp and the **Amplicon Must Include Region** defaults to the selection. |

*Table 23.1 Primer tab parameters – Amplification of Selected Region for BP Reaction dialog box*
### Primer Tab Parameters – Amplification of Selected Region for BP Reaction Dialog Box

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Number of Output Options</td>
<td>Restricted to 1 (i.e., 1 pair of primers and the associated product).</td>
</tr>
<tr>
<td>Salt Concentration</td>
<td>Common to all PCR tools. See page 248.</td>
</tr>
<tr>
<td>Probe Concentration</td>
<td>Ibid</td>
</tr>
<tr>
<td>dG Temperature</td>
<td>Ibid</td>
</tr>
<tr>
<td>Tm</td>
<td>Ibid</td>
</tr>
<tr>
<td>%GC</td>
<td>Ibid</td>
</tr>
<tr>
<td>Primer Length</td>
<td>Defaults to 18-25, Invitrogen’s recommendation for Gateway Primers</td>
</tr>
<tr>
<td>DNA/RNA radio button</td>
<td>Common to all PCR tools. See page 248.</td>
</tr>
<tr>
<td>Amplify on Complementary Strand</td>
<td>Check to amplify and express a sequence or feature which maps to the</td>
</tr>
<tr>
<td></td>
<td>complementary strand. When this is activated, the contents of the Sense</td>
</tr>
<tr>
<td></td>
<td>and Antisense browse-ins will be exchanged in the Primer Additions and</td>
</tr>
<tr>
<td></td>
<td>GGGG-attBx Extension boxes. For example, the default attB1 extension that</td>
</tr>
<tr>
<td></td>
<td>leads into the expressed sequence will be found in the Antisense text box</td>
</tr>
<tr>
<td></td>
<td>and the attB2 extension that exits from the sequence will be found in</td>
</tr>
<tr>
<td></td>
<td>the Sense text box.</td>
</tr>
<tr>
<td>Primer Additions</td>
<td>Enter nucleotides in the text box, or click the Browse button ( ) to add 5'</td>
</tr>
<tr>
<td></td>
<td>sequences to the template-specific region of the primer.</td>
</tr>
<tr>
<td></td>
<td>In the Choose Additions dialog boxes that open (Figure 23.4), there are</td>
</tr>
<tr>
<td></td>
<td>some recommended additions specific to the sense and antisense primers.</td>
</tr>
<tr>
<td></td>
<td>(Such sequences can assist in expression or purification when the template</td>
</tr>
<tr>
<td></td>
<td>sequence is transferred to an Expression Clone and expressed in a host</td>
</tr>
<tr>
<td></td>
<td>cell.) Click on the preferred addition, or click the Add button to name</td>
</tr>
<tr>
<td></td>
<td>and enter an additional sequence that can be added to the primer. Click OK</td>
</tr>
<tr>
<td></td>
<td>to enter your selection in the appropriate text box. Text box entries are</td>
</tr>
<tr>
<td></td>
<td>fully editable.</td>
</tr>
<tr>
<td></td>
<td>Note: If you type additions to the antisense window, enter the sequence in</td>
</tr>
<tr>
<td></td>
<td>5' ---&gt; 3' order, as you want it to appear on the complementary strand</td>
</tr>
<tr>
<td>Instruction</td>
<td>Click the Reading Frame and Stop Codons buttons to open context-sensitive</td>
</tr>
<tr>
<td></td>
<td>Online Help topics with tips for preserving a reading frame and inserting</td>
</tr>
<tr>
<td></td>
<td>stop codons into the primers.</td>
</tr>
<tr>
<td>Add GGGG-attBx 5’ Extensions</td>
<td>These text boxes must be filled in for a Gateway project. The defaults are</td>
</tr>
<tr>
<td></td>
<td>for single fragment cloning: attB1 for the sense primer and attB2 for the</td>
</tr>
<tr>
<td></td>
<td>antisense primer. Click the Browse buttons ( ) to replace the defaults</td>
</tr>
<tr>
<td></td>
<td>with other attB sequences for creating Entry Clones for multi-site cloning</td>
</tr>
<tr>
<td></td>
<td>projects. In the Choose attB Extension dialog boxes that open (Figure</td>
</tr>
<tr>
<td></td>
<td>23.5), click on the preferred extensions and click OK.</td>
</tr>
</tbody>
</table>

*Table 23.1 Primer tab parameters – Amplification of Selected Region for BP Reaction dialog box*
The attB extensions listed here are intended to ensure that the sequence of interest remains in-frame if it is ultimately cloned into a pDEST vector with N-terminal or C-terminal fusion partners.

In Multi-site Gateway experiments, 2 or 3 distinct elements are cloned into separate Entry Clones, each using a distinct pairing of attB extensions and specialized pDONR vectors. The Entry Clones are then mixed with a pDEST vector in an LR recombination (see below) to generate a single Expression Clone in which the elements are carried together. This can be useful in coordinating expression of disparate elements from one vector, or in constructing mini-genes. The following table summarizes the available lineups (as of summer 2005) for designing primers and matching the amplicons with their respective vectors.
Table 23.2 Multi-site Gateway recommendations for attB primer extensions and pDONR vectors (BP reactions), and pDEST vectors (LR reactions)

<table>
<thead>
<tr>
<th>No. of fragments</th>
<th>fragment order</th>
<th>attB extensions</th>
<th>pDONR vector</th>
<th>pDEST vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 fragment cloning</td>
<td>First fragment</td>
<td>SP: attB4 ASP: attB1</td>
<td>pDONR P4-P1R</td>
<td>pDEST R4-R2</td>
</tr>
<tr>
<td></td>
<td>Second fragment</td>
<td>SP: attB1 ASP: attB2</td>
<td>Any pDONR of type P1-P2</td>
<td></td>
</tr>
<tr>
<td>3 fragment cloning</td>
<td>First fragment</td>
<td>SP: attB4 ASP: attB1</td>
<td>pDONR P4-P1R</td>
<td>pDEST R4-R3</td>
</tr>
<tr>
<td></td>
<td>Second fragment</td>
<td>SP: attB1 ASP: attB2</td>
<td>Any pDONR of type P1-P2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Third fragment</td>
<td>SP: attB2 ASP: attB3</td>
<td>pDONR P2R-P3</td>
<td></td>
</tr>
</tbody>
</table>

The Load and Save buttons at the lower left of each tab are common to all PCR dialog parameter boxes. They allow you to save your PCR settings to a file and to load such settings files for subsequent analyses. See page 247 for more information.

In the context of this discussion, you might configure this dialog box to add a Lumio tag to a primer. By saving these settings, when you select other sequences for amplification, you could click the Load button and select the saved settings. The Lumio tag would automatically load in the Primer Additions text box and be added to the new primers when you click OK.

2. Once your PCR parameters are specified, click OK in the Amplification of Selected Region for BP Reaction dialog box.

PCR product and primers are calculated and displayed in the PCR Analysis Text Pane folder in the Molecule Display Window. Simultaneously, a dialog box opens (Figure 23.6), in which you can select to proceed directly to a BP reaction, and/or to add the primers to the Oligo Goal List to expedite ordering from Invitrogen’s custom Oligo Synthesis service. To bypass both options, click OK without selecting either checkbox; this will return you to the starting molecule so that you can examine the calculated PCR product and primers.

Figure 23.6 Choose to proceed directly to the cloning wizard, add the primers to your order list or examine the primers and products

- If you select Proceed to BP Reaction, the Create Entry Clone by BP Reaction dialog box (Figure 23.1) opens with the attB-containing PCR product(s) you generated appear-
ing automatically in the attB sequences pane. Continue as described in the previous section.

- If you elect to add the oligos to the Oligo List, you can order Oligos from Invitrogen at any time. See page 276 for more information.
- If you opt to exit here to examine the primers and products, you can pursue any of the following options and save the product and primers to continue at a later time:
  a. After examining the results, select the PCR product in the PCR Analysis Text Pane folder and select Recombine a PCR Product by BP from the shortcut menu. Continue with the Create Entry Clone by BP Reaction dialog box (Figure 23.1) as described in the previous section.
  b. Right click on the Product of Length (the PCR product) in the PCR Analysis Text Pane folder and select Save as Molecule in Database. At a later date, you can select the PCR product in the Database Explorer and Recombine a PCR Product to BP as described in the previous section. This option is useful if multiple PCR products are generated from different files, and you wish to store and submit them from the database in batch.
  c. Save the parent molecule along with all PCR products and primers in one molecule file. Choose File > Save As and in the Save As dialog box, select the Save as file tab and choose a location in which the files are to be saved. Later, to submit the saved PCR product to a BP reaction, open the file in a Molecule Display Window, select the PCR product in the PCR Analysis Text Pane folder and continue as described above.

Generate an Entry Clone by BP Recombination

The Create Entry Clone by BP Reaction dialog box opens with your amplicon in the attB Sequences pane. Continue with the BP reaction, beginning with step #2 on page 323. The BP recombination will transfer the amplified sequence, plus any additions you chose during primer design, into an Entry Clone.

Generate an Entry Clone by PCR Amplification and Ligation into a pENTR Vector

This method uses the Molecule Construction tool in Vector NTI and Invitrogen pENTR Vectors. pENTR Vectors are closed, circular DNA molecules with a multiple cloning site (MCS) and a ccdB gene flanked by attL1 and attL2. By amplifying a DNA sequence of interest with extensions that anneal to cut sites in the vector, the sequence can be inserted using Molecule Construction techniques. For more information on Molecule Construction, see Chapter 9 Tutorial: Molecule Construction and Chapter 22 Molecule Construction.

The generic workflow using pENTR vectors is as follows:

1. Open a Molecule Display Window for a selected pENTR vector and another for a DNA molecule with the sequence of interest. Use the menu command Window > Tile Vertical to conveniently view both files (Figure 23.7).
2. With the pENTR vector as the active window, define restriction site(s) on each side of the ccdB gene. Choose sites after and before attL1 and attL2, respectively, and avoid enzymes with internal sites in the sequence of interest. Choose List > Add Fragment to Goal List and follow the Fragment Wizard to add the fragment to the goal list. For directional cloning, choose sites that produce non-complementary overhangs.

3. With the DNA donor fragment as the active window, select the sequence of interest using standard Vector NTI selection techniques. With the sequence selected, choose from the menu Analyses > Primer Design > Amplify Selection.
   - In the Amplification dialog box that opens, set Before Max_BP and After Max_BP to 0, and Maximum Number of Output Options to 1.
   - Under More >>, using the Browse ( ) buttons, select 5' REN extensions for the sense and antisense primers that will anneal to the cut vector ends.
   - To amplify your selection, click OK.

4. In the PCR Analysis Text Pane folder, right-click on the PCR product and choose Save to Database and Create Window. In the New DNA/RNA Molecule dialog box that opens, enter the name of your new PCR product. Click OK. Select a subset to store the fragment in the DNA/RNA Molecule database and click OK.

5. In the Molecule Display Window displaying the new PCR product, select on the REN sites you designed into your primers (Figure 23.8), and choose List > Add Fragment to Goal List. Follow the Fragment Wizard and conclude by adding the fragment to the goal list.
6. From the **List** menu, choose **Show Lists**. From the List Dialog box menu, choose **Run**.

7. In the **Construct Molecule** dialog box, ensure that the **Circular** radio button is selected and enter a name for your Entry Clone. Choose a subset to store it in the Insert DNA/RNA Molecule(s) into Subset dialog box and click **OK**. The newly constructed Entry Clone will contain the DNA sequence of interest flanked by the **att**L sites. Save the new Entry Clone to the database.

8. Perform an LR recombination in Vector NTI (see below) and examine the resulting Expression Clone(s) by ORF analysis and translation to verify that the DNA sequence will be expressed as intended.

**Obtaining a Functional Entry Clone from Invitrogen**

Invitrogen offers collections of genomic sequences as **att**L-containing Source Clones (for example, the Ultimate ORF Collection). Check the Invitrogen website [http://www.invitrogen.com](http://www.invitrogen.com) or contact Invitrogen Technical Services.

**Performing an LR Recombination**

All the previously described methods in this chapter describe generating or obtaining an Entry Clone. In the Gateway LR reaction, one or more Entry Clones recombines with an **att**R-containing Destination Vector ((pDEST) to produce an Expression Clone. Vector NTI software performs LR recombination by accepting entries in several categories of projects (single-fragment, multi-site, and batch) in a single workspace, recombining them and automatically sorting and validating the entries. A preview screen summarizes all successful Expression Clones and allows autosaving of the results to the database. (For details on multi-site cloning, refer to the Support link on the Invitrogen website, [www.invitrogen.com](http://www.invitrogen.com).)

1. Open a Molecule Display Window and from the menu, choose **Cloning > Gateway Cloning > Create an Expression Clone by LR**. Alternatively, from the Database Explorer, select
one or more molecules, and choose **Gateway Cloning > Create an Expression Clone by LR** from the **DNA/RNA** or shortcut menu.

2. **First Screen:** In the Create Expression Clones by LR Reaction dialog box (Figure 23.9), use the **Add/Remove** buttons to browse in the pDEST vectors and Entry Clones, in the upper and lower panes respectively. If the tool was launched from an open file containing either type of molecule, those molecules will automatically appear in the appropriate pane. If invalid entries are browsed into either window, Vector NTI will issue a message and refuse the entry.

   **Tip:** The **attL** sites in an Entry Clone and **attR** sites in a pDEST vector must be compatible. For Entry Clones generated by Vector NTI software, the identity of **att** sites becomes part of the description annotation and can be displayed to help you select compatible molecules. When you click the **Add** button, the Choose Database DNA/RNA Molecules dialog box appears. Right click on the column header, and select **Columns**. In the Columns dialog box, select **Description** in the **Available Fields** panel to the left and move it to the **Displayed Fields** panel. You can click on the **Description** header to sort the molecules by description, or scroll through the list to find molecule(s) that contain the appropriate **att** sites.

   You can follow the same procedure in the **Database Explorer** to display **Descriptions** and thus recognize **att** sites generated by Vector NTI software for any Gateway Clone.

3. **Second screen:** An expandable preview page appears (Figure 23.10), summarizing the Expression Clones that can be generated from the available entries. Batch submissions (i.e. a single pDEST vector and multiple compatible Entry Clones) are summarized as separate Expression Clones. Check the **Save results to DB** checkbox to automatically save the Expression Clones to the Vector NTI database. Click **Finish**.
Note: If a pDEST vector and an Entry Clone share the same resistance markers, the recombination can proceed at the bench, but it will be difficult to select the final Expression Clone. If such Expression Clones can form, they will be noted on this page in orange. The simplest course in such cases is to recreate the Entry Clone using a pDONR vector with another marker, for example, pDONRzeo or pDONR207. Alternatively, consider using or creating a different pDEST vector.

The most likely situation for this to occur is when using new pDEST vectors that were not adequately annotated by the investigators who created them. (See Creating a New Destination Vector on page 335.) To minimize such occurrences, make sure that the new vectors have a unique marker and are annotated using permissible syntax, listed in Table 23.3. If your antibiotic resistance marker is not on this list, contact Invitrogen Technical Services to edit the appropriate configuration file. See Appendix A.

<table>
<thead>
<tr>
<th>Resistance Marker</th>
<th>Permissible Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>Ampicillin, ampicillin, Ampicillin, AmpR, AP(R), bla</td>
</tr>
<tr>
<td>Blasticidin</td>
<td>blasticidin, Blasticidin, Bsd, Bsd(R)</td>
</tr>
<tr>
<td>Gentamicin</td>
<td>Gentamycin, Gentamicin, gentamycin, gentamicin, gentimicin, gentimycin, Gm(R), GmR</td>
</tr>
</tbody>
</table>

Table 23.3 Permitted Syntax for Antibiotic Resistance Markers
To automatically save the results to the Vector NTI database, check the **Save results to DB** checkbox. You are offered the option to open the new clone in a Molecule Display Window. Click **Finish**.

If the **Save results to DB** option was checked, the Expression Clones are saved but do not open automatically in Molecule Display Windows. If the **Save to the DB** option was not checked, Expression Clones open in Molecule Display Windows as temporary molecules that can be saved in the Database.

Examine the resulting Expression Clone(s) by ORF analyses and translation to verify that the DNA sequence is expressed as intended. See the following section.

### Verifying Expression Product
To verify that the DNA sequence is expressed as intended, examine the resulting Expression Clone(s) by ORF and translation analyses (Figure 23.11).

<table>
<thead>
<tr>
<th>Resistance Marker</th>
<th>Permissible Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hygromycin</td>
<td>Hygromycin B</td>
</tr>
<tr>
<td></td>
<td>hygromycin B</td>
</tr>
<tr>
<td></td>
<td>Hygromycin B(R)</td>
</tr>
<tr>
<td></td>
<td>hygromycin B(R)</td>
</tr>
<tr>
<td></td>
<td>Hygromycin(R)</td>
</tr>
<tr>
<td></td>
<td>hygromycin(R)</td>
</tr>
<tr>
<td></td>
<td>Hygromycin</td>
</tr>
<tr>
<td></td>
<td>hygro</td>
</tr>
<tr>
<td></td>
<td>Hygro(R)</td>
</tr>
<tr>
<td></td>
<td>hygro(R)</td>
</tr>
<tr>
<td></td>
<td>hyg</td>
</tr>
<tr>
<td></td>
<td>hph</td>
</tr>
<tr>
<td></td>
<td>hyg(R)</td>
</tr>
<tr>
<td></td>
<td>hph(R)</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>Kanamycin</td>
</tr>
<tr>
<td></td>
<td>kanamycin</td>
</tr>
<tr>
<td></td>
<td>KmR</td>
</tr>
<tr>
<td></td>
<td>KN(R)</td>
</tr>
<tr>
<td></td>
<td>Kn(R)</td>
</tr>
<tr>
<td></td>
<td>Kan(R)</td>
</tr>
<tr>
<td></td>
<td>Kan-Tn5</td>
</tr>
<tr>
<td>Neomycin</td>
<td>neomycin</td>
</tr>
<tr>
<td></td>
<td>Neomycin</td>
</tr>
<tr>
<td></td>
<td>Neo(R)</td>
</tr>
<tr>
<td></td>
<td>NeoR</td>
</tr>
<tr>
<td>Spectinomycin</td>
<td>spectinomycin</td>
</tr>
<tr>
<td></td>
<td>Spectinomycin</td>
</tr>
<tr>
<td></td>
<td>SpnR</td>
</tr>
<tr>
<td></td>
<td>Spn(R)</td>
</tr>
<tr>
<td>Zeocin</td>
<td>zeocin</td>
</tr>
<tr>
<td></td>
<td>Zeocin</td>
</tr>
<tr>
<td></td>
<td>ZeoR</td>
</tr>
<tr>
<td></td>
<td>Zeo(R)</td>
</tr>
<tr>
<td></td>
<td>Sh ble</td>
</tr>
</tbody>
</table>

Table 23.3 Permitted Syntax for Antibiotic Resistance Markers
Figure 23.11 ORF and translation analysis on Expression Clone verifies that an expressed sequence is in-frame with a N-terminal fusion partner (a GST tag).

1. For ORF Analysis, select **Analyses > ORF**.

2. To verify translation, select **Analyses > Translation > Into New Protein > <translation option>**. If the DNA sequence was cloned into a Destination Vector containing N- or C-terminus fusion partners, make sure that they and sequence of interest are contained within a single ORF. Check that the reading frame translates the fusion partners and the sequence of interest correctly, and that no premature stop codons are encountered. If the reading frame is disrupted, it may be necessary to return to primer design for creating Entry Clones again. For detailed tips, see the Reading Frame Online Help topic available in the Primer tab.

3. In molecules created by the Gateway cloning tools, a folder **Design Description** appears in the Text Pane. It contains a brief description of the protocol and can be copied and pasted into a word document for reference during the actual experiment.

### Creating a New Destination Vector

Invitrogen offers a wide selection of Destination Vectors for expressing a sequence of interest in a variety of hosts and conditions. However, you may want to convert your own vectors into a Gateway Destination Vector in order to exploit unique promoters, fusion tags, etc. To assist in this process, Invitrogen offers sets of blunt-ended Reading Frame Cassettes for introducing **attR1** and **attR2** sites, and the requisite selectable marker genes, into any vector to convert it into a Destination Vector.

The set consists of three cassettes, each of which defines a different reading frame from vector features (tags, cleavage sites, etc.) to **attR** sites (and ultimately, to the insert sequences cloned in by LR recombination).

1. Browse the local database to your vector of choice and open it in a Molecule Display Window.
2. Choose **Cloning > Gateway Cloning > Create a Novel Destination Vector** from the drop-down menu.

![Create a new Gateway pDEST vector – first screen](image1)

**Figure 23.12 Create a new Gateway pDEST vector – first screen**

3. **First screen** (Figure 23.12):
   - Following the instructions, **CLICK** (and **SHIFT-CLICK**) on the vector feature map to select 1 (or 2) cut sites on the vector for inserting the cassette. If the digestions produce overhangs, choose a method to blunt the vector ends, then **Click Next** to continue.
     - **Fill-in overhangs**
     - **Trim back overhangs** (If any selected site generates 3’ overhangs, only this option will be offered).

![Create a new Gateway pDEST vector – second screen](image2)

**Figure 23.13 Create a new Gateway pDEST vector – second screen**

4. **Second screen** (Figure 23.13):
   - From the drop-down menu, choose a **Reading Frame Cassette**: rfa_verB, rfb_verB or rfc.1_verB. The cassettes differ in length by 1 bp, so that one of them will produce a pDEST vector with an unbroken reading frame between fusion partners in your vector (if any) and the attR sites introduced with the cassette.

**Note**: Some combinations of cut sites and cassettes generate in-frame stop codons during ligation, which would prevent expression of a fusion protein between vector sequences and the sequence coded from inserts. Vector NTI will issue a message if such a combination is encountered, and recommend alternative actions if expression of a fusion protein is being considered. If the original vector lacks fusion partners, these instructions can be ignored.

5. Choose to generate one or two new Destination Vector(s) with the cassette in direct, reverse or both orientation(s).
6. Click **Finish**. Inspect the new vector(s) and design restriction, sequencing or PCR screens with which to test transformants for cassette orientation.

**Introduction to TOPO Cloning**

TOPO cloning takes advantage of the activity of topoisomerase I, one molecule being covalently bound to each end of a linearized vector in a proprietary process. Inserts up to 3 kb in size are readily accepted. In a 5 minute bench-top reaction, incoming DNA fragments with 5'-OH ends are unwound by the enzyme and ligated to the vector end, in a process that releases energy and liberates the enzyme. The vectors can be grouped into 3 categories, based on the nature of their ends:

- **Zero-Blunt** vectors have 2 blunt ends and can accept blunt ended DNA fragments, including amplicons produced by a proofreading polymerase. Inserts are cloned in both orientations.

- **T-A** vectors have 2 ends with 3'-T overhangs. They can accept products of PCR amplification with a *taq* polymerase, whose terminal transferase activity adds 3'-A overhangs to the amplicon. Inserts are cloned in both orientations.

- **In directional** vectors one terminal is blunt ended and the other has a 5'-GGTG overhang on the bottom strand. PCR products are generated with a 5'-CACC extension on one end and this strand when unwound is preferentially annealed to the vector overhang. More than 90% of the clones are in the correct orientation and the time spent in screening colonies is thereby reduced.

The presence of topoisomerase enzyme also helps protect vector ends from degradation, particularly from contaminating nucleases that may be present in ligase preparations. Moreover, the avoidance of restriction site cutback for cloning PCR products means that internal cleavage sites are not a problem.

**Generate a Clone by TOPO cloning**

Any linear, double-stranded DNA sequence of interest may be cloned into a TOPO vector using VNTI NTI Advance. Circular molecules will not be accepted. In addition, linear sequences with 3'-A overhangs, the products of PCR amplification with a *taq* polymerase, may be cloned into TOPO-TA vectors. Such *taq*-generated molecules can be generated *in silico* using the Vector NTI Advance PCR tool **Amplify selection to use in TOPO reaction** (see below).

1. Open the sequence in a Molecule Display Window, and from the menu, choose **Cloning > TOPO Cloning > Launch TOPO Wizard**.

   Alternatively, in the Database Explorer, select one or more such molecules and from the DNA/RNA or shortcut menu choose **Launch TOPO Wizard**.
2. **First screen** (Figure 23.14): Use the Add/Remove buttons to browse in one or more sequences and TOPO vectors, in the upper and lower panes, respectively. If the tool is launched from open windows displaying either type of molecule, those molecules automatically appear in the appropriate pane. Use the information in the Type fields to ensure compatibility of insert sequence and TOPO vector. Click **Next**.

(If the insert sequence by chance has ≥75% identity to CACC at both 5’-terminals, either naturally or due to PCR (see below), the type field will report the insert as double directional. These inserts are accepted by TOPO directional vectors, but the directional cloning ability is lost. In such cases, consider adding a cap to one of the primers, so that the insert will be cloned in the intended orientation.)

3. **Second screen** (Figure 23.15): An expandable preview page appears, summarizing valid clones that can be generated from the entries displayed on Screen 1. Note that vectors are regarded as “unused” if they are incompatible with the insert sequences (T-A in this case). Note also that cloning with bi-directional TOPO vectors will produce 2 clones, with the insert sequence in both orientations. Batch submissions from the first screen (e.g., a single vector and multiple compatible sequences) are summarized as separate clones.

4. To automatically save the results to the Vector NTI database, check the **Save results to DB** checkbox. You are offered the option to open the new clone in a Molecule Display Window. Click **Finish**.
In molecules created by the TOPO cloning tools, a folder **Design Description** appears in the Text Pane. It contains a brief description of the protocol and can be copied and pasted into a word document for reference during the actual experiment.

5. If the **Save to the DB** option was not checked, clones will open in Molecule Display Windows as temporary molecules that can be saved in the Database (Figure 23.16).

![Figure 23.16 Clones generated with the selected sequence cloned in opposite directions](image)

**Generate a Clone by PCR Amplification and TOPO Cloning**

This strategy is a highly flexible, rapid process for generating clones by amplifying a sequence of interest with PCR and combining the amplicon with a TOPO vector. The sequence may be on either strand of any type of DNA molecule, linear or circular, since the first step will generate a linear PCR product in the desired orientation. You can choose between amplifying for combination with Zero-Blunt, directional or T-A TOPO vectors. For the latter, Vector NTI simulates the action of *taq* polymerase, including the addition of 3'-A overhangs to the amplicon by terminal transferase activity. In addition:

- You can add additional sequences to the primers to assist in protein expression and purification.
- You can open online help topics for assistance in the placement of stop codons and maintaining reading frames.
- If reading frame errors are detected, you can return and make the appropriate edits.
- You can add primer recommendations to an Oligo Goal List from which orders can be submitted to Invitrogen’s custom Oligo Synthesis service.
1. Open a Molecule Display Window for a DNA molecule containing the sequence of interest (Figure 23.17). Select the target region on the molecule and choose Cloning > TOPO Cloning > Amplify selection to use in TOPO reaction. (These commands are disabled if an appropriate target is not defined.) This opens the Amplification of Selected Region for TOPO cloning dialog box (Figure 23.18).
2. This dialog box allows you to search for convenient sense and antisense primers according to your specifications and to check the uniqueness of primers. Most of the tabs and parameters in this dialog box are common to those found in the Find PCR Primers dialog box discussed above with regard to Gateway cloning (Table 23.1). The difference on the Primer tab is described here. For more information on setting PCR parameters in this dialog box, see page 241.

The options on the Primer tab (Figure 23.18) allow you to generate a single pair of primers that amplifies but does not extend beyond the selected sequence. It also allows you to tailor the amplicon termini for specific types of TOPO vectors.

a. In the box marked Cloning termini, choose Blunt to generate an amplicon with 2 blunt ends. These will be the exact boundaries of the selection. These amplicons are best used with Zero-Blunt TOPO vectors.

b. Choose T-A to generate an amplicon as would be produced by amplification with a taq polymerase. The primers in such a case will anneal to the exact boundaries of the selection. However, terminal transferase activity in the enzyme will add 3’-A overhangs to each end of the amplicon (Figure 23.19).

c. Choose directional for cloning in TOPO directional vectors, e.g. pENTR-D/TOPO. The amplicon will be generated using primers, one of which includes a 5’-CACC extension. The choice of primer is governed by the checkbox Amplify on complementary strand. In the example here, the small T antigen maps to the complementary strand of SV40. In order to orient it for cloning and expression the checkbox will be on. The resulting amplicon is shown in Figure 23.20. Note the CACC extension 5’ to the start codon on the complementary strand.

![Figure 23.19 SV40 small T amplified with computed primers and taq polymerase. Note the 3’-A overhangs exterior to the PCR primers.](image-url)
Figure 23.20 SV40 small T amplified for reaction with a TOPO directional vector. Note the 5’-CACC extension on the complementary strand.

As described above, the **Load** and **Save** buttons at the lower left of each tab are common to all PCR dialog boxes. They allow you to save your PCR settings to a file and load the settings file in for subsequent analyses. See page 240 for more information.

3. Once PCR parameters are specified, click **OK** in the Amplification of Selected Region for TOPO cloning dialog box.

4. PCR product and primers are calculated, and display in the **PCR Analysis Text Pane** folder in the Molecule Display Window. Simultaneously, a dialog box opens (Figure 23.21), in which you can select to proceed directly to a TOPO reaction, and/or to add the primers to the Oligo Goal List to expedite ordering from Invitrogen’s custom Oligo Synthesis service. To bypass both options, click **OK** without selecting either checkbox; this will return you to the starting molecule so that you can examine the calculated PCR product and primers.

Figure 23.21 Choose to proceed directly to the cloning wizard, add the primers to your order list or examine the primers and products.
If you select **Proceed to cloning wizard**, the Create TOPO Clone dialog box (Figure 23.14) opens with the PCR product(s) you generated appearing automatically in the inserts pane. Continue as described in the previous section.

If you elect to add the oligos to the Oligo List, you can order the custom Oligos from Invitrogen at any time. See page 266 for more information.

If you opt to exit here to examine the primers and products, you can pursue any of the following options and save the product and primers to continue at a later time:

a. After examining the results, select the PCR product in the **PCR Analysis Text Pane** folder and select **Clone PCR Product by TOPO** from the shortcut menu. Continue with the Create TOPO Clone dialog box (Figure 23.14) as described in the previous section.

b. Right click on the **Product of Length** (the PCR product) in the **PCR Analysis Text Pane** folder and select **Save as Molecule in Database**. At a later date, you can select the PCR product in the Database Explorer and **Launch TOPO wizard** as described in the previous section. This option is useful if multiple PCR products are generated from different files, and you wish to store and submit them from the database in batch.

c. Save the parent molecule along with all PCR products and primers in one molecule file. Choose **File > Save As** and in the Save As dialog box, select the **Save as file** tab and choose a location in which the files are to be saved. Later, to submit the saved PCR product to a TOPO reaction, open the file in a Molecule Display Window, select the PCR product in the **PCR Analysis Text Pane** folder and continue as described above.

5. Generate a TOPO Clone. The Create TOPO Clone dialog box opens with your amplicon in the attB Sequences pane. Continue with the TOPO reaction, beginning with step #2 on page 315. The amplified sequence, plus any additions you chose during primer design, will be cloned into the chosen vector (Figure 23.22).

---

**Figure 23.22** Small T cloned into Gateway-enabled TOPO vector pENTR-D/TOPO. Note that the amplicon is reoriented so that direction of expression is in the attL1→attL2 direction. This clone is ready for recombination with a Gateway pDEST vector in an LR reaction.
High Throughput Operations

The simplicity of Gateway and TOPO cloning easily lend themselves to automation and high-throughput operations. This final section describes procedures that can rapidly generate and store high numbers of primers, amplicons and clones.

High Throughput PCR

1. Open one or more Molecule Display Windows containing the sequences of interest. If more than one window is opened, consider Tile or Cascade from the Window menu to arrange them on screen for convenience.

2. For each open window, follow these steps
   a. Select the entire sequence (CTRL-A or Select All from the Edit menu) and add to the Multiplex PCR List (List menu > Multiplex PCR List).

   b. The Lists window opens on the Multiplex PCR tab (Figure 23.23). Click Add on the Lists window menu and in the ensuing Add Fragment to Multiplex PCR Analysis dialog, choose either Add Whole Selection or Add Features Inside Selection.

   c. If you choose Add Whole Selection and click OK the entire sequence from the respective Display Window is entered in the multiplex list.

   d. If you choose Add Features Inside Selection, use the Browse button (+) to specify the feature types to amplify (e.g. CDS, exons, etc.). You can specify individual instances of the feature type by name. Or accept the default Add all features of this type. Click OK and the Add Fragment to Multiplex PCR Analysis dialog will list the named features, or Any <feature type>. The Add Fragment to Multiplex PCR Analysis dialog can be used to mix and match different feature types on one or several molecules.
e. Click OK and the Lists window will display a row <molecule name> Features <feature type/name> inside Selection <molecule length>

f. Repeat these steps for all molecules open in Molecule Display Windows.

3. When the list is complete, click either Run GW or Run TOPO from the Lists window menu. The Multiplex PCR Analysis dialog opens for the respective cloning operation. For this illustration, we will use Run GW (Figure 23.24).

![Multiplex PCR Analysis dialog for generating Gateway amplicons](image)

Figure 23.24 Multiplex PCR Analysis dialog for generating Gateway amplicons

4. The parameters are configured to generate sense and antisense primers that amplify but do not extend beyond the selected features. As in single-fragment cloning, sequence additions can be added/typed to assist in expression or purification. The attB-extensions appearing in the Add GGGG-attBx 5’ extensions boxes will be added to respective sense and antisense primers.

Ideally, all features for amplification should be on the same strand of the open files, most often the direct strand. If some features are on the opposite strand, you should consider whether the cloning procedure downstream is directional. (TOPO Zero-Blunt and TOPO T-A are not and both orientations will be produced anyway) If cloning is directional (directional TOPO or Gateway), you can repeat the analysis, and in the list setup, specify the features on the opposite strand by name. Then check on the Amplify on Complementary Strand checkbox.

5. In addition, these parameters are configured to attempt amplification of the selected molecules/features without high regard to standard multiplex criteria.

   - Multiplex fold minimum = 10
   - Amplicon Length Difference Min = 0
   - Maximize clusters

Most strikingly, there is no attempt to ensure that amplicons in a cluster will be a minimum size difference from one another (Amplicon Length Difference Min = 0). It is assumed that the PCR reactions computed here will be executed in plate format, one reaction per well, rather than true multiplex format. If this is not the case in your experiment, adjust these parameters accordingly. Finally, the default configuration will record results in a Report file (Results tab > Generate Report).
6. Click OK. When PCR computations are complete, a report **Preview: Multiplex PCR Amplification** is generated (Figure 23.25), which summarizes the features that can be amplified and the primers computed for the PCR reactions. Under the configuration described above, most reactions will be grouped in one or a few clusters. Click **Save Report** to save the report to a file where it may be read in any text editor. Then click **Close Report** and check the preferred options from **Save PCR products as DNA molecules to DB**, **Save Primers as Oligos to DB** and **Add Primers to Oligo List**. Both “clustered” and unclustered amplicons, and their Gateway primers, are saved to subbases in their respective databases (Figure 23.26).
Figure 23.26 Over 200 annotated features in HUMBRCA1 were amplified and deposited in a single subbase of the DNA/RNA database. Four features that were successfully amplified but not “clustered” were deposited in a separate subset (“Cluster_unclusterized”)

High throughput cloning

One or more entries in the Explorer DNA/RNA database can be selected by standard operations (CLICK, SHIFT + CLICK, CTRL + CLICK). When this is done, the Gateway or TOPO cloning tools can be launched from the DNA/RNA or shortcut menu. Let’s continue with the example using the Gateway amplicons generated above.

1. Click on the Cluster 1 subbase in the DNA/RNA database and select all of its contents.
2. From the DNA/RNA or shortcut menu, choose Gateway Cloning > Recombine an Expr Clone by BP (Figure 23.27).

3. The first screen of the BP cloning tool opens, with all the amplicons listed in the attB sequences pane.
4. Continue as with BP for single fragment cloning. Add a pDONR vector, and click **Next**. You have the option of directly opening all resulting clones in Molecule Display Windows. Click **Finish**. The clones will be stored in a subset of your choosing.

5. To generate Expression Clones in batch, repeat steps 1 through 4., but choose **Gateway Cloning > Create an Expression Clone by LR** from the **DNA/RNA** or shortcut menu.
Introduction

This chapter describes the creation of DNA molecules through Design.

In Design, you can describe the list of fragments in very general or very specific terms. You can also describe preferences for techniques used in creating the molecule, such as how fragments may be isolated from their parent molecules, how termini may be modified, etc. In contrast to the Construction technique described in Chapter 22, in Design Vector NTI does the rest, taking advantage of the best possible restriction sites and genetic engineering techniques to engineer the molecule.

This chapter describes:

- The overall procedure for designing a DNA molecule
- How to describe a goal molecule using Display windows, Fragment Wizard, and the Molecule Editor
- How to start the design process
- How to set your preferences for design techniques
- What Vector NTI actually does during the design process, and how you can control the process
- Discussion of the design plans Vector NTI generates.

Creating a New DNA Molecule Using Design

For creating a molecule by Design, the Molecule Display window must be in Molecule Edit mode, the default mode when the window is opened.

The process of creating a new molecule with Vector NTI’s Design capabilities generally follows these steps in an active Graphics Pane:

1. Describe your goal molecule by defining its recipient and donor fragments and placing them in the Goal Molecule Definition List in the proper order using Fragment Wizard.
2. Open the Lists dialog box, loaded with the component fragments for your Goal Molecule.

3. Open the Design Molecule dialog box, enter general information for the new molecule and initiate the Design.

4. When the Design Parameters dialog box appears, set the appropriate parameters and your design preferences.

5. Start the design process. Vector NTI searches for optimum Design options. If Design is successful, your molecule is entered into the database and opens in its new Molecule Display window.

6. In the Text Pane of the display window containing the goal molecule, open the Design Description folder to inspect the design plan Vector NTI has generated.

If you are not satisfied with the design plan, you can choose to redesign the molecule by changing the description of the goal molecule or using different parameters.

**Describing the Goal Molecule**

In Design, you must define one *recipient fragment*, listed first in the Goal Molecule Definition List, and one or more *donor fragments*.

The simplest way to define fragments for Design is to use the Fragment Wizard and the Graphics Pane to visually define fragments. You can then use the Fragment Editor to fine-tune fragment descriptions. Fragments can be defined by hand in the Design Molecule dialog box, but it is much more tedious and much less effective than using Fragment Wizard. Refer to Chapter 22 for a detailed description of Fragment Wizard. Here we will describe how to use it to define design recipient and design donor fragments.

**Defining the Recipient Using Fragment Wizard**

- The recipient fragment of the goal molecule remains unaffected during the Design process. All nucleotides within the recipient fragment you defined are included in the goal molecule. This means that any polylinker sites used for cloning must be outside the recipient fragment.
- The recipient must be listed as the first fragment in the Goal Molecule Definition List. This will be covered later in the chapter.
- The recipient must come from a circular molecule stored in the Vector NTI database.

To define a recipient fragment, open the Display window for the recipient molecule. If you have several display windows open concurrently, you can display the source molecules concurrently by selecting *Window > Tile Vertically* or make sure the source molecule for your fragment is the topmost window. Activate its Graphics Pane and click the *Add Fragment To Goal List* button ( ), or select *List > Add Fragment to Goal List*, opening The Fragment Wizard dialog box (Figure 24.1).

![Figure 24.1 Fragment Wizard dialog box](image)

The Fragment Wizard dialog box content changes to display the entry fields and descriptions corresponding to the current step in the fragment definition process.
At any point in the process, you can choose to **Cancel**, go **Back** a screen, go to the **Next** screen, or **Finish** the process using the appropriate buttons. **(Finish** goes directly to the last step, skipping all the remaining steps, if any. **Fragment Wizard** goes through all the steps you skipped and supplies the missing information using default values and the coordinates of the current selection in the Display window.)

*In the first “screen”:*

Select the kind of fragment you wish to define from three choices: Since this chapter covers Design, choose the **Design Recipient Fragment** option and click the **Next** button to go to the second screen of **Fragment Wizard**:

*In the second screen (Figure 24.2):*

To specify the 5' terminus of the fragment, select one of the following options. **Vector NTI will look for good cloning sites in the region upstream from the position you specify.**

- Specify a position on the molecule map (in one of the two following ways)
  1. Enter the coordinate in the text box or click in the Graphics Pane of the display window. In the latter case, the Set to a Position button will be selected automatically and the caret position is displayed in the position field of the dialog box (and on the Status Bar). To reposition the caret, use the right and left arrow keys.

**Tip:** If, when you click in the Graphics Pane, the entire molecule is selected (as indicated by the wire-frame), click in the Sequence Pane to deselect, and begin again.

- Specify the start of the molecule by selecting the Start of Molecule option in the dialog box.

- Specify a restriction site.

To set the terminus to a restriction site, click on the site or its label in the Graphics Pane. **If selected**, **Vector NTI must use that site in cloning. For Design, you may be setting an unrealistic limitation. If the site is not convenient, for example, requiring complicated partial digestion, the cloning may be very difficult.**

**Note:** If the **Fragment Wizard** screen limits your view of the molecule map, left click on Wizard’s blue title bar, and drag it out of the way.

![Fragment Wizard](image)

*Figure 24.2 Specifying the 5' terminus of the fragment*

When the 5' terminus is set, click the **Next** button to proceed.

*In the third screen (Figure 24.3):*

This screen appears only if you specified that the 5' terminus should be set to a palindromic restriction site. If that was not your choice, move on the description of the fourth screen.

You are asked to specify whether the site should be saved or lost in the designed molecule. Your options are: **Save Site**, **Lose Site** or **Don't Care**. If you do not specify that site is to be
saved or lost, its fate will be unpredictable. Choose the desired option and click the Next button to advance to the next screen of the dialog box.

![FragmentWizard](Image)

**Figure 24.3** Specifying whether the site should be saved or lost in the designed molecule

**In the fourth screen:**

Specify the 3’ terminus of the fragment. Vector NTI will look for good cloning sites in the region following the position you specify. Options are:

- Specify a position on the molecule map (in one of the two following ways)
  1. Enter the position into the text box or SHIFT + CLICK in the Graphics Pane and adjust the caret position using SHIFT-arrow key combinations described in Chapter 3.
  2. Specify the end of the molecule
- Specify a restriction site.
  SHIFT + CLICK on the site or its label in the Graphics Pane. If selected, Vector NTI must use that site in cloning. For Design, you may be setting an unrealistic limitation. If the site is not convenient, for example, requiring complicated partial digestion, the cloning may be very difficult.

**Tip:** The SHIFT key must be held down for all actions in the Graphics Pane for setting the 3’ end. If you neglect to do this, your actions will change the 5’ terminus and Fragment Wizard will revert to screen 2.

If you specified that the 3’ terminus is set at a palindromic restriction site, you can go to the next screen and specify that the site should be saved or lost in the designed molecule. Choose the desired option and click the Finish button to complete the definition of the fragment.

If you did not specify a palindromic restriction site, the Next button is disabled; click the Finish button, signifying completion of fragment definition.

![NewFragment](Image)

**Figure 24.4** Adding the fragment to the Goal Molecule Definition List

Review the description of the selected fragment in the New Fragment message box. If you made a mistake, click the Cancel button to return to the Fragment Wizard. Otherwise, click the Add to List button (Figure 24.4), adding the fragment to the Goal Molecule Definition List and closing the Fragment Wizard.

**Notes about Design:**
- If you later edit the recipient fragment in the List/Design Molecule dialog box, any biochemical operations you order to be performed on the termini of the recipient fragment will be ignored. During the Design process, Vector NTI chooses all methods of terminus modification itself.

- The recipient fragment always has direct orientation. The Inverted box in the recipient’s Fragment Editor dialog is disabled during the Design process.

**Defining the Donors Using Fragment Wizard**

Donors are fragments that Vector NTI inserts into the recipient fragment (vector) to produce the goal molecule. Donor fragments in Design must carry functional signals into the recipient.

When performing Design, Vector NTI requires all fragments except the first in the Goal Molecule Definition List to be donor fragments. Donor fragments can come from circular or linear molecules stored in the database.

To define a donor fragment, open the display window for the donor molecule. *(If you have several display windows open concurrently, the source molecule for your fragment should be the topmost window.)* Activate its Graphics Pane and click the **Add Fragment To Goal List** button \( \text{Add Fragment To Goal List} \) on the Window toolbar or select **List > Add Fragment to Goal List**, opening The Fragment Wizard dialog box.

*In the first screen:*

Select the Design Donor Fragment option and click the **Next** button to go to the second screen of the Fragment Wizard.

*In the second screen (Figure 24.5):*

- To select the desired functional signals to be carried on the donor fragment into the recipient, click on the signal or its label in the Graphics Pane. To select more than one signal, hold down the **SHIFT + CLICK**. **Note:** If the Fragment Wizard screen limits your view of the molecule map, left click on Wizard’s blue title bar, and drag it out of the way.

- To deselect a selected signal, hold down **SHIFT + CLICK** again.

*Figure 24.5 Selecting the desired functional signals to be carried on the donor fragment into the recipient*

The names of the signals you selected are displayed in the Fragment Wizard dialog box.

If the donor molecule is circular and you have chosen more than one signal, Fragment Wizard lets you choose the place of the “gap” between signals where Vector NTI will look for a convenient place to cut the donor molecule. To move the gap clockwise, click the right arrow button on the Fragment Wizard; counterclockwise, click the left arrow button. To illustrate this, select more than one signal and experiment. When the desired signals are selected, click the **Next** button to go to the next screen.
In the third screen (Figure 24.6):

The next screen for specifying the 5’ terminus for design offers two selections:

- **Leave Terminus Undefined** if you want Vector NTI to choose restriction sites for you.
- **Use Specific Site** to use a restriction site to cut the signals from the donor molecule. Fragment Wizard prompts you to hold down the SHIFT key and click on the restriction site or its label in the Graphics Pane. The site label is selected and the name of the site and its position are shown in the Fragment Wizard dialog box. If selected in this way, Vector NTI is forced to use that site in cloning. If the site is not convenient, for example, requiring complicated partial digestion, you will be responsible for the difficulty of the cloning situation.

![Fragment Wizard](image)

*Figure 24.6 Specifying the 5’ terminus for design*

Click the **Next** button to go to the next screen.

In the fourth screen (Figure 24.7):

If you have chosen the **Use Specific Site** option, this screen lets you specify the 3’ terminus for design. The required actions are the same as for the 5’ terminus screen described above.

If you have chosen the **Leave Terminus Undefined** option, this screen asks you to specify the maximum size of the 5’ flank region (if you came from the 5’ terminus definition screen) or 3’ flank region.

![Fragment Wizard](image)

*Figure 24.7 Specifying the 3’ terminus for design, or specifying the maximum size of the 5’ flank region*

- **Use Flank Region No Larger Than ... bps** limits the region within which cloning sites may be sought (thus limiting the size of the resulting flank region). At the prompt, enter either the maximum length of the flank region in the dialog’s text box or move the appropriate end of the selection in the Graphics Pane into the gap beyond the edges of the selected functional signals. In this case, the selection in the Graphics Pane shows a maximum possible flank region. You can use all available techniques for extending a selection in the Graphics Pane: drag the appropriate end of the selection, SHIFT-CLICK on the desired position, etc. For details on Selection Techniques, refer to Chapter 3.
**Do Not Limit the Size of a Flank Region** allows Vector NTI to search all available space for restriction sites. When you are finished with the description of both termini of the donor fragment, the **Next** button becomes disabled. Click the **Finish** button to complete the definition of the donor fragment. The New Fragment message box is displayed with the description of the selected donor fragment. If you made a mistake, click the **Cancel** button to return to the Fragment Wizard dialog. Otherwise, click the **Add to List** button. The fragment is added to the Goal Molecule Definition List and the Fragment Wizard dialog box is closed.

The **Goal Molecule Definition List (Goal List)**

The Goal Molecule Definition List (Figure 24.8) (in the **Fragments** tab at the bottom of the Lists dialog box) is a list of fragments to be combined to create a new molecule. In addition to adding fragments to the Goal Molecule Definition List using the Fragment Wizard, you can add them directly from the Lists dialog box.

To see the Goal Molecule Definition List, click the **Open Goal List** button or choose **List > Molecule Goal List**. In the Lists dialog box (**Fragments** tab) that opens, the Fragments display box lists the fragments currently in the Goal Molecule Definition List.

![Figure 24.8 Goal Molecule Definition List](image)

**Note:** The first fragment listed is considered the “recipient fragment.” If a fragment is listed in incorrect order, highlight it and click the **Up** or **Down** buttons to reposition it.

**Starting the Design**

Although the goal molecule component fragments are ready for processing by Vector NTI’s Design mode, remember that VNTI decides on the optimum Design fragment termini and cloning strategies.

To begin the design of your new molecule, open the Lists dialog box by clicking on the **Open Goal List** button or selecting **List > Molecule Goal List**.

Open the Design Molecule dialog box by clicking the **Run** button on the Lists dialog box.

**Note:** The Construct/Design Molecule dialog box has two operational modes—a mode for molecule construction and mode for molecule design. Because you have defined fragments in this chapter for Design, in this mode, the dialog box title is the Design Molecule. See Chapter 22 for details on the Construct Mode.

Enter or select the dialog box fields as follows:

- **Name**
- **Circular/Linear** - the form of the molecule.
- **DNA/RNA** - molecule type
- Click the General Info button to enter molecule description information:
  - **Description** - notes about the molecule
  - **Extra-Chromosome Replication** - transformation systems
  - **Replicon Type** - molecule’s replicon type
  - **Keywords** - keywords useful for database search. Refer to page 171 for details.
  - Click OK to return to the Design dialog box.

All of the **General Info** data will appear in the General Description folder when the molecule is loaded into a Display window.

- **Recipient’s Start** - start of the new molecule is as close as possible to the start of the “recipient” component fragment (the first fragment listed of the Component Fragments).
- **Position ...** - specify a specific nucleotide of a specific fragment as the starting nucleotide of the new molecule. (*By default, the first nucleotide of the first fragment.*)

### The Component Fragments Field

Designed molecules consist of one or more component fragments. Use the Fragments field in the Lists dialog box to verify or enter this information.

You may edit, delete, and change the order of the fragments in the Component Fragments field, and you may add new fragments of five types. These operations are described in Chapter 22.

It is essential in molecule design that the recipient fragment is first in the Fragments list, followed by the donor fragments. If you need to change this order, highlight the fragment(s) to be moved and click the **Up** or **Down** buttons until they are in the correct order.

### Defining Design Fragments in Fragment Editor

An alternative to using Fragment Wizard for defining design fragments is to enter them manually into the **Fragments** tab of the Lists dialog. To do this, click the **Add** button and choose **Add <Donor/Recipient> Fragment > with Fragment Editor** to open the Fragment of Molecule dialog box (Figure 24.9).

**Figure 24.9 Fragment of Molecule dialog box**

Options in the Fragment Editor dialog box are summarized as follows:

- **Design recipient/donor fragment**
A donor fragment created with a Display window always has a defined parent molecule name.

Change the orientation of a donor fragment by checking the Inverted box.

Left/Right Terminus buttons open the Edit Terminus dialog box (Figure 22.16). The way you defined the termini initially governs which sections are enabled in this dialog box. For more information on modifying termini, refer to Chapter 22.

In Design, all donor fragments carry functional signals called descriptors into the recipient. To add or remove descriptors from a donor fragment, use the For Design Donors Only section of this dialog box. To add a descriptor, click the Add button, opening the Fragment Descriptor dialog box (Figure 24.10):

![Fragment Descriptor dialog box](image)

To define a descriptor for the donor fragment, choose the type of signal and enter the name of the signal, if necessary. If the signal is on the complementary strand, check the Complementary box. (If you do not enter a signal name, the descriptor is added to the donor fragment and is treated as any signal of the specified type.)

**Note:** The Fragment Descriptor dialog box is the only way to add a descriptor containing any signal of a specified type. Such a descriptor cannot be added from a display window.

Click OK to return to the Fragment of Molecule dialog box (Figure 24.9).

- To remove a descriptor from a donor fragment’s description, highlight it in the Descriptor(s) box and click the Delete button.
- If after leaving the Display window you would like to add flank regions to a donor fragment, or change flank regions set through a Display window, you can use the Maximum Flank Region boxes.
- Subset of Parents – If you prefer to simply identify the feature to be added to the new molecule, you can let Vector NTI decide which specific molecule of the subset is used as the source for the functional signal. In the Subset of Parents drop-down menu, select the subset where the parent molecules might be sought.

After entering and verifying the data in the Lists dialog box, you are ready to proceed with the design process. Click the Run button to launch the Design Molecule dialog box and then click the Design button.

Vector NTI performs a preliminary save of the current molecule state. The molecule’s name, general information, fragment list, etc. are saved and can be recalled. The system then performs a preliminary analysis of the design possibilities, and opens the Design Parameters dialog box, described below.

If you start a design for a goal molecule with a name that causes a conflict with an existing set of design options, Vector NTI informs you of the conflict and ask permission to delete the existing...
options. If you do not want to do this, choose a different name for the new goal molecule or rename old design options.

**Design Parameters**

After the Design is initiated, in the interim dialog box that opens, select a subset for storing Design results. The Design Parameters dialog box (Figure 24.11) then opens where you can set the Design parameters the program uses to design the molecule.

![Design Parameters dialog box](image_url)

Figure 24.11 Design Parameters dialog box

Options in the Design Parameters dialog box are summarized as follows:

- **REN Subsets** displays the list of restriction endonucleases that Vector NTI will consider in the current Design task.

- **Transformation System(s) and Extra-Chromosome Replication** specifies the capabilities for the current Design process.

- **Dephosphorylation Permitted** allows dephosphorylation to be used as a pre-selection method.

- **Try to Lose Junction Sites** directs the loss of junction sites if possible without increasing cloning complexity. Lost sites are then used for efficient pre-selection. This box is a request, not a command, and Vector NTI may save the site if that results in a more convenient design option. *If, in the Fragment Wizard, you specified that sites be saved or lost in the description of the recipient, they will automatically be saved or lost regardless of whether this box is checked.*

- **Fragment List Gives Order of Cloning Steps** inserts the first donor into the recipient in the first cloning step, the second donor in the second cloning step, etc. *If this box is unchecked, the order of cloning steps is based on the most convenient restriction sites.*

- **Fragment List Gives Order in Goal Molecule** places the donor fragments in the goal molecule in the same order in which they are listed in the goal molecule description. *If this box is unchecked, the placement order is based on the most convenient restriction sites.*

- **Delete Region** specifies the number of nucleotides to be deleted from the recipient molecule during cloning. *This may be disabled, depending on how the fragments were defined.*

- **Phage Size May Be Changed By** is enabled if the recipient molecule is a phage, which may lose its functionality due to length changes. If you want to preserve phage functionality in the goal molecule, it may be useful to prevent the goal molecule from
being much different in length than the recipient molecule. Enter the maximum change in length (%) for phages in the field. *If this field is enabled and specified, then the lengths of the goal molecule and all intermediate recombinants produced by the Design process are kept within the specified length limits.*

- **Preferences** opens the Design Preferences dialog box. See next section.
- **PCR Primers** – Vector NTI can use PCR amplification to isolate donor fragments. The system also will propose primers for PCR amplification of cloned fragments in all molecules produced by the Design process, including the goal molecule.
  - Click the **Parameters** or **Quality** buttons under PCR Primers to set the desired parameters or Quality Specifics of PCR primers. Refer to page 253 for the details of the dialog boxes that open with these commands.
  - The primers used will satisfy the conditions set by the user in the PCR Analysis dialog box.

When all design parameters are set as you desire, click the **Start Design** button.

### Design Preferences

Although much of the Design process is decided by Vector NTI, you can select preferred techniques for designing the molecules. Click the **Preferences** button in the Design Parameters dialog box (Figure 24.11) to open the Design Preferences dialog box (Figure 24.12):

![Design Preferences dialog box](image)

In this dialog box, you can set preferences in the categories of fragment isolation, fragment ligation, and terminus modification.

Depending upon your Design results, you may find it useful to have Vector NTI design the same goal molecule several times, changing the design preferences each time. This gives you several alternatives for constructing your molecule.

The parameters for this dialog box are as follows:

- **Fragment isolation**: Vector NTI isolates fragments for cloning using convenient restriction sites whenever possible. If no convenient restriction sites are present, partial digestion or PCR can be used to isolate fragments. Check the isolation methods you will permit:
  - Partial Digestion (Simple): Partial Digestion with 1 site inside required fragment and 3 sites total.
- **Partial Digestion (Medium):** Partial Digestion with 1 site inside required fragment and not more than 6 sites total.
- **Partial Digestion (Complicated):** Partial Digestion with more than 1 identical site inside required fragment.
- **PCR**
- **Ligation:** Check the junction types you will permit Vector NTI to use in designing molecules. **Cohesive...Cohesive** junction means both ligated fragments have two cohesive ends, each longer than 1 nucleotide; **Cohesive...Short junction** means both ligated fragments have one cohesive and one short (exactly 1 nucleotide) end, etc.
- **Terminus modification:** Check the methods you will permit to make termini compatible.

**Other Preferences**

- **Separation in Gel:** Specify the minimum relative difference in length (%) for separating fragments in gel.
- **Complicated Partial Digestion:** Specify the maximum number of restriction sites inside the required fragment and the total number of sites on the molecule that are acceptable for fragment isolation using complicated partial digestion.

**Priorities**

**What’s Better** – rate priorities for the techniques and ligation types by moving them in the corresponding lists. To reposition any item, select it and click the **Up** and **Down** buttons. Items near the top of the lists have higher priority than items lower in the lists.

**Advanced Design Preferences**

Click the **Advanced** button in the Design Preferences dialog box to open the Internal Design Parameters dialog box (Figure 24.13) to fine-tune Vector NTI’s heuristic search for molecule design options. **Default settings are given in italics.**

![Internal Design Parameters dialog box](image)

**Depth of Backtracking:** Specify the number of steps Vector NTI should retrace in its search algorithm to design a molecule if it hits a dead-end in finding a molecule according to the defined parameters. **Default value is 5**

**Min. Rating for Acceptance:** Specify the minimum acceptable rating to immediately accept an option and end the current search step. Vector NTI rates design options on an integer scale of 1 to 8. The higher the number, the longer Vector NTI continues searching for a better construct according to the defined parameters. **Default value is 3**

**Min. Rating to Avoid Next PD Level:** If Vector NTI cannot achieve a minimum rating for acceptance (see previous point), it may try the next more complicated level of PD (partial digestion.) For instance, if you have listed Simple Digest first (under What’s Better in Fig-
Molecule Design

Chapter 24

When you click the Start Design button in the Design Parameters dialog box, Vector NTI begins the Design process. An interim “Designing...” dialog box appears, showing you the progress of the design and allowing you to break and accept at certain points, or cancel the design entirely. The design process consists of several stages:

1. Vector NTI collects all the information it needs to perform the design. The message “Collecting necessary data...” is displayed on the screen.

2. Vector NTI creates restriction maps for all molecules, which may be involved in the design. If many maps need to be created (for example, if you chose a large subset of molecules as potential donors) and if many restriction endonucleases are involved, Vector NTI may take a relatively long time to create all the maps.

3. Vector NTI begins to perform cloning steps. In each step, one donor fragment is cloned into the recipient.

4. When each cloning step is complete, Vector NTI constructs the best recombinant it obtained for that step, models cloning analysis after transformation, searches for important restriction sites in the recombinant, etc.

5. If a cloning step produces the goal molecule, that is, if no donors remain to be cloned, Vector NTI finishes the Design task and enters the goal molecule into the database.

6. If donor fragments remain, the system moves on to the next cloning step.

7. When the Design process is complete, all intermediate recombinants (if any) are stored in a new subset in the database, called name_$Intermediate, where name is the goal molecule’s name. A Molecule Display window opens only for the final molecule.

If you start a Design task for a goal molecule with a name that causes a conflict with an existing set of intermediate recombinants, Vector NTI informs you of the conflict and asks permission to delete the existing intermediates and turn the existing designed molecule into a basic molecule. If you do not want to do this, choose a different name for the new goal molecule.

The recombinants are named as follows: name_$step number_variant number, where name is the goal molecule’s name, step number is the number of the cloning step, and variant number is the number of the variant of this step, if more than one variant has been generated.
From a special subset created for the Design result molecules, you can open, examine and/or delete the intermediate recombinants as desired. *If you delete intermediate recombinants, the goal molecule becomes a basic molecule.*

From the moment Vector NTI begins making restriction maps, the Cancel button is active. Press it if you want to abort the design at any point. If you cancel the design, the information in the Design Molecule dialog box remains unchanged.

During each cloning step, if at least one acceptable design option has been found, the Accept Now button is active. Click it to stop further search and accept the best option already found for the current cloning step. You may wish to turn off the **Immediate Break When There Are Enough Backtracking Options** check box in the Advanced Design Preferences, and control the depth of the heuristic search for each cloning step with the Accept Now button.

### Returning to the Pre-Design State

When Vector NTI designs molecules, it saves the original goal description (list of recipient and donor fragments) in the database along with all other information about the new molecule. After the molecule has been designed, you can return it to its original state.

This is useful if you want to try several different designs based on the same goal molecule description. Design the molecule one way and inspect the results in a Molecule Display window. If you are not satisfied with the results, change the design parameters or design preferences, and try again until you are satisfied.

Select **File > Molecule Operations > Advanced DNA/RNA > Design**. An intercept dialog box opens reminding you that you will be overwriting the new molecule with the original fragments. Click Yes to continue with the new Design. The Lists/Design Molecule dialog box opens again loaded with the original Design component fragments you selected.

If you cannot achieve a satisfactory design by changing the design parameters and preferences, try changing the description of the goal molecule. Change the molecule names each time so you will have a record of your attempts.

If you still cannot achieve a satisfactory design, try changing the internal design parameters to influence Vector NTI’s heuristic searches for design options.

### The Design Description Folder

The Molecule Display window for the final new molecule automatically opens upon the conclusion of Design. At a later time, you can open a Molecule Display window for newly designed intermediate molecules stored in the database. In either case, the Text Pane of the display window contains a Design Description folder with subfolders for each cloning step used to create the molecule. In each cloning step, one donor fragment is cloned into the recipient.

#### Cloning Step Subfolders

Each cloning step subfolder of the Design Description folder contains the following information:

- **Result Molecule**: Name of the recombinant obtained in this cloning step. If this is the final cloning step, the name of the recombinant is the same as the name of the goal molecule. If this is an intermediate recombinant, the name is formed according to the rules given previously in this chapter.

- **Recipient**
  - The name of the molecule used as the recipient
  - The restriction sites used as the left and right termini
  - If and how the termini were modified
If partial digestion was used in this cloning step, the number of sites inside the required fragment and the total number of sites on the molecule

- **Donor**
  - The molecule used as the donor
  - Whether the fragment was inserted in complementary orientation,
  - The restriction sites used as the left and right termini of the donor
  - If and how the termini were modified
  - If partial digestion was used in this cloning step, the number of sites inside the required fragment and the total number on the molecule
  - The actual flank regions on each end of the donor fragment, that is, the distance in nucleotides between the first or last nucleotide in the fragment and the appropriate boundary of the nearest functional signal
  - If the donor fragment was obtained through PCR amplification, the PCR primers used are displayed in the 5’-3’ direction. If a restriction site was attached to the 5’ end of the primer, that site is also shown here.

- **Ligation Junction types**: cohesive or blunt.
- **Lost restriction sites on the recipient** indicates sites lost during cloning
- **Donor fragment orientation** indicates if the ligation conditions prevent parasitic donor fragment orientations.
- **Preselection**: The recommended pre-selection method
  - If one or both recipient sites are lost after ligation, they will be recommended for effective pre-selection.
  - If sites are not lost, the system will recommend dephosphorylation (if that was permitted in the Design Parameters dialog box).

The ability to perform pre-selection influences the choice of methods of clone analysis after transformation.

- **Transformation system**: The recommended transformation system for clone analysis. This information is useful if the user allowed more than one transformation system to be used. Vector NTI considers the size of recombinants, their replicon types, and extrachromosome replication capabilities.
- **Recommended method(s) for clone analysis**: Recommended methods for clone analysis after transformation. Vector NTI proposes the most reasonable set of four possible methods: Screening, Colony Hybridization, Restriction Analysis, and PCR amplification of the cloned fragment. These recommended methods are based on a number of conditions: ligation type, full or partial digestion, pre-selection, etc.
- **Recommendations for restriction analysis**: The recommended restriction sites give distinguishable gel patterns for the result molecule (recipient with insertion) and the recipient without insertion.

The following information is given even if these methods were not recommended for clone analysis:

- **Recommended oligonucleotide for colony hybridization**: The oligonucleotide will successfully hybridize with the inserted fragment.
- **Primers for PCR**: allow amplification of the region closest to the boundaries of the cloned fragment.
- If the ligation conditions permit parasitic donor orientations, restriction sites for restriction analysis are recommended to identify the clone with required donor fragment orientation.
- Important restriction sites for the result molecule are listed and categorized: sites missing on the molecule, unique sites located within the cloned fragment, and unique sites located outside the cloned fragment.

**Redesigning a Molecule**

You can redesign a previously designed molecule by right-clicking on the designed molecule in the Database Explorer window and choosing **Redesign** from the shortcut menu. This launches the Design Molecule dialog box with all the parameters used for the original design of that molecule. The Redesign feature is useful for designing molecules that are similar to a previously designed molecule and for redesigning molecules that have failed a previous design attempt (such molecules appear with a question mark icon to the left of their names in the Database Explorer window).
Introduction

Vector NTI's Gel Display windows are graphical tools designed for supporting your electrophoresis work.

In a Gel Display window you can:

- Add gel markers and samples to gel lanes.
- View general electrophoresis data or specific fragments in Text Pane folders.
- Assign colors and line formats to individual gel fragments.
- "Run" a gel in increments or animation for a specific length of time.
- Enlarge or reduce a gel for optimum visualization.
- Move gel samples or markers to and from the Gel Sample List, a data clipboard or display windows.

Opening a Gel Display Window

Gel Display windows can be opened in two ways: by opening a new display window or by opening a previously saved Gel Document file.
Opening a New Gel Display Window

To open a new Gel Display window, click the **New Gel** button ( ) on the Main Toolbar, or choose **Gel > Create New**. The Gel Setup dialog box (Figure 25.1) opens for setting gel parameters:

- **Electrophoresis Profile**: Select from preset settings or previously saved settings
  - **Save Settings As…** Click to name and save new settings specified in this dialog box.
  - **Remove Saved Setting** Select an entry in the profile list and click this button to remove it
- **Electrophoresis Type**
  - **Agarose gel with constant electric field**
  - **Polyacrylamide (Constant Field)**
  - **Pulse Field** (Agarose gel)
- **Electrophoresis Parameters** *Default settings are in italics.*
  - **gel concentration**: 1.0%
  - **electric field**: 2.5 V/cm
  - **gel length**: 15.0 cm
  - **pulse time** (for pulse field EP) 2 sec
- **View Parameters**
  - **Time increment** (time span that each gel modeling iteration adds or subtracts from the gel run)
  - **Animation speed** (time increments in minutes added to the gel run for each second of real time during animation.)
  - **Separation distance** (the minimum distance by which fragments must be separated in order for you to distinguish them in the gel.)
- **Buffer** All possible buffers are represented by two types.
  - **Tris-Acetate (TAE)**
  - **Tris-Borate (TBE)**

Click **OK** to close the settings box and open a Gel Display window. *See the Electrophoresis Profile bulleted point to the save the settings.*
To modify the settings of a Gel Display window at any time, click the **Gel and Display Setup** button ( ), or select **Edit > Gel and Display Setup**. The Gel Setup dialog box opens where you can change data for the open display window.

**Opening a Gel Document File**

Gel Display window contents can be stored in a Gel Document file. To open an existing Gel Document file, select **Gel > Open Document**. In the standard File Open dialog box, browse for the directory or folder and file name of the Gel Document to be opened. Click **Open**.

**Elements of a Gel Display Window**

A Gel Display window is divided into two sections, or panes: a Text Pane and a Gel Pane, divided by a split bar. When the Gel Display window is first opened, it remains empty until you create your gel.

The Gel Display window toolbar has two different variants, depending on whether the active pane is the Text Pane or the Gel Pane. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

**Gel Samples and Gel Markers**

Vector NTI gels can contain two types of data:

1. Gel samples: fragments to be analyzed in a Gel Display window. They are created for short-term use. *If saved to the database, they become gel markers.*

2. Gel markers: sets of fragments of known lengths used as standards for measuring migration speed. Gel markers are stored and loaded from the Vector NTI database.

**Creating and Adding a Gel Sample to a Gel**

To create a gel sample for one gel lane, click the **Create Sample** button ( ) on the Window toolbar or choose **Gel > Create Gel Sample**, opening the Create Gel Sample dialog box (Figure 25.2):

![Figure 25.2 Create Gel Sample dialog box](image)

- **Source Molecules** – In the left column, select a DNA/RNA Molecule subset and one or more molecules from that subset.
Source Enzymes – In the center column, select an Enzyme subset and one or more enzymes from that subset.

Selected for Gel Sample – The right column of the dialog box shows molecules and enzymes currently selected for creating the sample.

Removing Entries – To remove a molecule or enzyme from the Selected for Gel Sample column, deselect it in the source column or double-click on its name in the Selected for Gel Sample column.

Performing Partial Digestions – Notice that each enzyme name in the selected column has a checkbox to its left. Checking the box indicates to Vector NTI that you wish to perform a partial digestion for the enzyme.

Sample Name/Description – Enter a sample name and description in the text box at the bottom of the dialog box. The gel data is identified in Text Pane folders by the name and description you enter in these boxes.

When the gel sample is defined, click one of the action buttons at the top of the dialog box:

Add to Gel – Adds the sample to one lane in the active Gel Display window. Select new molecules and enzymes for each lane and click Add to Gel. If there is no active Gel Display window, this button is disabled.

Add to Gel Sample List – Adds the sample to the Gel Sample List. The Gel Sample List dialog box is described below.

Save as Gel Marker – Opens the New Gel Marker dialog box (Figure 25.3) allowing you to save the sample to the Database as a gel marker (see the next section).

Note: In the Create Gel Sample dialog box, you must select a molecule. To open the New Marker dialog box without having a molecule already selected, choose Gel > Create Gel Marker.

Click the OK button in the New Marker dialog box to save the sample as a gel marker and to return to the Create Gel Sample dialog box. To add the new marker to the gel, see Adding a Gel Marker to a Gel on page 369.

New Gel Marker Dialog Box

To create a new gel marker, select Gel > Create Gel Marker. The New Gel Marker dialog box opens to the General tab. Enter a name for the new marker. Once you have entered the name, select the Gel Marker tab.
The Fragments list box lists by length all fragments making up the current sample.

- To add a fragment to the sample, enter its length in bp in the New Fragment box and click the < Add button.
- To remove a fragment from the sample, select it in the Fragments box and click the > Delete button.

Tip: If by chance you press Enter instead of the < Add key, thus closing the dialog box before you are finished, go to the Gel Markers table in the Database Explorer. Double click on the new gel marker you were creating. It opens in an Edit Gel Marker dialog box where you can continue the process.

On the Gel Marker tab you can also edit the sample’s description, shown in the Description text box. The User Fields, Comments, and Keywords tabs are the same as for other kinds of database objects; they are described in Chapter 20.

Click the OK button to save the sample to the Database as a gel marker.

Adding a Gel Marker to a Gel

To add a gel marker to a gel, click the Add Marker Lane button ( ) on the Gel Toolbar. In the Choose Database Gel Marker dialog box (Figure 25.4), select the gel marker and click OK, adding it to a gel lane. (If you have created your own gel marker, you will see it listed in this dialog box.)

Figure 25.4 Choose Database Gel Marker dialog box

The Gel Sample List

The Gel Sample List (Figure 25.5) is a data clipboard for mixing gel sample fragments taken from various sources within Vector NTI. This is an alternative way to prepare gel markers.

Figure 25.5 Gel Sample List
Viewing the Gel Sample List

To open the Gel Sample List, click the **Open Gel Sample List** button ( ) on the Main Toolbar or choose **List > Gel Sample List**.

To remove a fragment from the Gel Sample list, select the fragment and click the **Remove** button. To remove multiple fragments, select all the desired fragments click the **Remove** button.

To save the Gel Sample List as a gel marker, click the **Marker** button. This opens the Gel Marker Editor, described previously.

Adding Fragments to the Gel Sample List

To add a fragment to the Gel Sample List, select it in the Gel Pane by CLICKING + DRAGGING over its bands with the mouse. Then click the **Add to Gel Sample List** button ( ), or choose **List > Add Selection to Gel Sample List**.

If the Gel Sample List is empty, the fragments you have selected will be added to the list. If the Gel Sample List is not empty, you will be given the option of adding the currently selected fragments to the existing set, or of having the selected fragments override the existing set.

Loading Fragments from the Gel Sample List

To load fragments from the Gel Sample List to the gel, click the **Add Lane from Gel Sample List** button ( ), or choose the **Edit > New > Add Lane from Gel Sample List**. All of the fragments currently contained in the Gel Sample List are added to a new lane of the active Gel Display window.

Panes in a Gel Display Window

Once a gel has been created, the contents of the two panes become apparent (Figure 25.6):

![Text Pane and Gel Pane of a Gel Display window when a gel has been created](image)

Text Pane folders contain a listing of electrophoresis gel data pertinent to the gel shown in the Display window, and information about the samples loaded into the gel’s lanes.

The Gel Pane shows a graphical depiction of an electrophoresis gel you have just loaded.
To switch back and forth between the panes, use the **Switch Pane** buttons on the Gel Pane Toolbar ( ), ( ). The display window panes can be manipulated using the scroll bars and split bars, as explained in detail in Chapter 3.

**Gel Display Window Text Pane**

The Gel Display window’s Text Pane contains one or more folders that can be opened, closed, expanded, collapsed, etc. as described in Chapter 3.

The Gel Display window Text Pane always has a General Description folder, containing data about the gel itself, including parameters set in the Gel Setup dialog box. The folder also shows the separation distance, time increment, and animation speed for this gel.

The Text Pane also contains a folder for each occupied lane. Folders can contain:

- a line giving the length range of the fragments in the lane
- a description of the contents of that lane
- subfolders (gel sample), labeled by fragment length, for each fragment in the lane. For gel samples created and not yet saved, the subfolders contain a description of the source of the fragment, for example, "Source: ColEl; HaeII(5419) - HaeII(1012). The numbers refer to the REN site position on the source molecule map.
- subfolders (gel marker), while giving the fragment lengths, do not provide fragment sources

**Modifying Fragment Properties**

You can use the Text Pane to set the color and pattern with which a fragment is displayed in the Gel Pane. This is convenient for tracking the progress of a particular fragment as you run your simulated gel.

To set a fragment’s display color, select the specific fragment’s folder in the Text Pane. Select **Edit > Sample Fragment Properties** or **Sample Fragment Properties** from the shortcut menu. In the Format Sample Fragment dialog box, choose a pattern and color for the fragment, and click **OK**. The fragment is displayed in the Gel Pane with the appropriate pattern and color. You may not be able to see it until you run your gel.

To simply view the properties for an entire lane, select the lane number by clicking on its well in the Graphics Pane and choose **Edit > Lane [#n] Properties**, or select the corresponding option from the shortcut menu or press ALT + ENTER.

**The Gel Display Window Gel Pane**

The Gel Display window Gel Pane displays a simulated gel, the lanes of the gel running vertically in the pane. Each lane is numbered at the top of the lane in a box called a “well.” As the gel progresses, a size scale runs down the left side of the Gel Pane, showing fragment sizes in bp.

To activate the Gel Pane, click on the toggle switch on the Gel Pane Toolbar.

**Step Forward, Step Back and Animation**

The Time box on the Display window toolbar ( ) tracks the time the gel “has been running”, in the form **hours:minutes**. To set a new time, click this box to activate it, enter the new time and press ENTER.

Click the **Step Forward** ( ) and **Step Back** ( ) buttons to advance or reverse the gel one increment of time per click.
Click the **Animate** button ( ) to start the gel running continuously. To stop the animation, click the **Animate** button again or the **Step Forward** or **Step Back** button.

(The time increment and the animation speed were set in the Gel Setup dialog box [Figure 25.1])

**Gel Lanes**

A Gel Display window can contain any number of lanes. You can select a lane by clicking on its well.

To change the lane width, move the mouse cursor to the division between any two wells where the cursor changes to a double-ended arrow ( ). Drag the cursor left or right. Since all of a gel's lanes are displayed with the same width, if you change the width of one lane, all of the other lanes change as well.

To perform other operations on lanes, place the cursor within the lane's well and open a shortcut menu, with these options:

- **Lane [#] Properties** displays the properties for the selected lane.
- **Move Left** moves a lane one slot to the left. (Or click on the well and select **Edit > Move > Move Lane Left**.)
- **Move Right** moves a lane one slot to the right. (Or click on the well and select **Edit > Move > Move Lane Right**.)
- **Delete** deletes the selected lane. (Or click on the well and choosing **Edit > Delete [#] Lane**.)
- **Select All** selects all fragments in the lane.
- **Add to Gel Sample List** adds all selected fragments to the Gel Sample List.
- **Save as Marker** launches the New Gel Marker dialog box, allowing you to save the selected fragments as a gel marker.

**Selecting Fragments**

To select fragments in the Gel Pane, drag the cursor that becomes a horizontal I-beam ( ) vertically across the fragments to be selected. All fragments in the highlighted selection zone are selected.

There are two reasons for selecting fragments:

1. Selected fragments can be copied to the Gel Sample List or saved as a gel marker.
2. If the selected fragments are too close to distinguish, VNTI can calculate how long the gel would have to run to separate them (Separation Time).

Copying fragments to the Gel Sample List and calculating separation times are described below.

**Zooming In and Zooming Out**

The Gel Pane Zoom buttons, which operate as in other Vector NTI windows, allow you to enlarge or reduce the gel lane display.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

The **True-Scale View** button ( ) is specific to the Gel Display window: Click this button to display the gel in an approximation of true scale. *When you select this display option, Vector NTI looks up the pixel size (dot pitch) of your monitor and displays the gel based on that value.*
Separation Distance

After the Gel Pane has been enlarged several times, gray bars representing separation distances appear around each fragment. Separation Distance is the minimum distance by which fragments must be separated in order to be distinguishable.

The separation distance is specified in the Gel Setup dialog box (Figure 25.1).

To eliminate the gray zones, click the Zoom Out button several times.

Calculating Separation Times

To calculate a separation time, select any number of fragments, at least two of whose gray bars are touching. Click the Calculate Separation Time button ( ) on the Window toolbar or choose Analyses > Calculate Separation Time. The minimum time necessary to separate the selected fragments is displayed. The gel running time is then set to that value.

Note: To calculate Separation Time, at least two fragments must be selected.

Vector NTI prompts you if any one of the selected fragments would reach the end of the gel before all selected fragments have separated.

Modifying Fragment Properties

Note: To modify fragment properties to more easily track fragment migration, refer to the section Modifying Fragment Properties under Gel Display Window Text Pane earlier in this chapter.

Copying Gel Display Window Data

To copy information from a Gel Display window to the Clipboard or a file, use Vector NTI's Camera feature. General directions for using this feature are described in Chapter 3. Directions specific for the Gel Display window are given here.

When the pane to be copied is activated, click on the Camera button ( ). Select in the dialog box that opens the range and destination of the copied material. The dialog box that opens from the Gel Pane is unlike most of the other Camera dialog boxes, but is specific for the data to be copied (Figure 25.7).

The copy options are described as follows:

- **Lanes**: Include selected lanes.
- **Comb**: Include lane numbers (above selected lanes).
- **Scale**: Include scale (to the left of selected lanes).
- **Gray Zones**: Include separation thresholds (gray rectangles around fragments).
- **Borders**: Include borders between lanes
- **Colors**: Use fragment colors

Copied Gel Pane information is in Windows Metafile format.

### Printing from Gel Display Windows

Printing from Gel Display windows is done one pane at a time. To print a pane, select it and click the **Print Active Pane** button ( ) on the main toolbar or choose **Molecule > Print**. In the standard Print dialog box, click **OK**.

Further print details (Page Setup and Print Preview) are given in Chapter 3.
CHAP 26
REPORT GENERATION

Introduction

Vector NTI Report Generation Wizard is a convenient tool for customizing different types of reports of DNA/RNA or protein molecule data. This chapter provides step-by-step instruction for creating both reports and report templates using the Report Generation Wizard.

Report results are presented in a hierarchal structure, similar to the Text Pane of the Molecule Display window. The Print Preview feature reflects the actual appearance of the generated report to be printed. The Vector NTI Camera feature copies selected information to the Clipboard or file in text format, preserving the formatting and enabling easy data exchange with other applications.

Two types of reports can be generated:

- **General**: This report represents general molecule(s) information, such as general description, standard fields, Vector NTI specific fields, user-defined fields, authoring information, comments, Feature maps, restriction maps, sequence, components and fragments and all analyses performed for selected objects.

- **Relation**: The second report represents the parent/descendant relationship between molecules as well as construction information.

Vector NTI report templates enable you to store the settings for generated reports and retrieve them to use at a later time instead of default settings.

Starting the Report Generation Wizard

To generate a report on a molecule open the molecule in a Molecule Display Window and choose **File > Reports > <report type>** on the menu bar. Two of the options initiate Report Generation Wizard. The third allows you to work with report templates. Each of the three report options are discussed in this chapter.
The Report Generation Process:

Step 1: Opening the Report Generation Wizard Dialog Box

Selecting General or Relation (Reports) opens the same Report Generation Wizard dialog box, shown in . For both types of reports, follow the outlined steps to collect the required information.

![Figure 26.1 Report Generation Wizard dialog box](image)

Use the **Next >** and **< Back** buttons at the bottom of the windows to proceed through the steps. Cancel the process at any time by pressing the **Cancel** button, or at any step press the **Finish** button to generate the report according to the current settings.

The first step offers two options:

- **New Report** lets you create a new report using the default settings. This initializes the options with default settings.
- **Begin a Report based on an existing template**. This initializes the options with previously saved settings.

Click **Next >** to proceed with either a new General or Relation Report.

To generate a report from an existing template without altering any settings, click **Finish** after selecting the desired template.

Step 2: Selecting General Options

In the Set General Report options window (Figure 26.2), check the informational fields with a left click or press the space bar. Due to the hierarchical organization of the fields, if you include the parent item, all of the descendants are automatically included. Excluding descendants automatically excludes their parent item as well. Grayed items indicate that the item’s state cannot be changed. This notation usually indicates either mandatory fields or information that could be included in the report, but are unavailable currently because they were not generated prior to starting the generation wizard.
Note: The more fields that are selected for the report, the longer the report generation takes.

![Select General Report Options window](image-url)

**Figure 26.2 Select General Report Options window**

- **General description**: This group of options consists of general information fields such as name, molecule length, form, and molecule type and nature of the sequence. Molecule name, type and length fields are mandatory and cannot be excluded from the report information.
- **Standard fields**: Standard fields include object data found routinely with all Vector NTI molecules and other objects. This includes for example, keywords, original source database, accession numbers, sequence source and identifiers in Vector NTI.
- **User-defined fields**: Check to include all user-defined fields in the report
- **Authors**: Specify the original and/or current authoring information
- **References**: Enter any references associated with the molecule
- **Comments**: Add any comments associated with the molecule
- **Feature map**: Specify molecule Feature map information. If this field is checked, the Wizard adds an additional step for adjusting Feature map information.
- **Restriction map (DNA/RNA)**: Check this item to include a restriction map in the report. The output can be represented either as a tree (similar to the restriction map information in the Text Pane of the molecule display window), or as a table.
- **Motifs (DNA/RNA)**: Check this item to include motifs in the report. The output can be represented either as a tree (similar to the restriction map information in the Text Pane of the molecule display window), or as a table.
- **ORF (DNA/RNA)**: This outlines the ORF information that is present.
- **Analysis**: As analysis information is not stored in the database, these options are available only if you start the Report Generation Wizard from a Molecule Display Window. Your report can include only those analyses that have been performed on a molecule and are shown as available options.
**DNA/RNA molecules**

Supported options are:

- PCR analyses
- Hybridization probes
- Sequencing primers
- Mutagenesis
- Restriction fragments

**Proteins**

Supported options are:

- Protein properties
- Acid content
- Sequence: Check this option to include the molecule sequence.

- **Construction:** This option generates construction information for complex molecules, including components and fragments of other molecules.

Click **Next >** to proceed to the next window. *If you have not selected Feature map, you will go directly to.*

**Step 3: Adjusting Feature Map Information**

![Set Feature Map Options dialog box](image)

This step is necessary if you choose to include Feature map information. Select from the Set Feature Map Options window (Figure 26.3) list the features to be separately included or excluded from your report (according to the molecule type). All available feature types are organized into the same tree structure that is in the Display Setup Profile window.
Step 4: Report Preview Window

The steps for either a General or Relation report end at a Preview dialog box (Figure 26.4). Clicking Finish at any step in the process also brings you directly to this dialog box. The results, represented in a hierarchical structure, allow you to hide information you think unnecessary by closing the appropriate branches.

Scroll through the Preview window with the scroll bars. ToolTips appear if you pause the cursor over a partially visible item. Open shortcut menus on selected items to reveal commands for expanding and collapsing folders and subfolders in various combinations.

This window lets you review the report and go Back if there is something you would like to change, or Save the report as a Template. If you are satisfied with the report, proceed with Print Preview or Print the report (details below) or use the Camera feature to copy the report to the clipboard or to a file. For details about using the Camera/Copy feature, refer to Chapter 3.

Print Preview and Print

To preview results before printing, click Print Preview. Navigate through the pages using the Next Page and Prev Page buttons. Use the Zoom buttons to enlarge or reduce the image. To print the results, click the Print button.
To modify something, click the Close button, returning you to the report. You can go back several steps and change the generation options, or you can expand or collapse some branches to customize your report presentation.

![Print Preview](image)

*Figure 26.5 Preview results before printing using the Print Preview feature*

### Saving Report Templates

To save the report format as a template, click the **Save Template** button, then enter a name and click **OK** on the Save Report Template window to save it. Templates are saved separately for report and molecule types. For example, when generating a relationship report for DNA/RNA molecules, the Report Wizard offers you templates saved for the same type of report and molecule type.

![Save Report Template](image)

*Figure 26.6 Save Report Template window*

### Relation Report

The Relation report includes information on parents, construction and descendants of the molecules involved.

- The parents of a molecule are the molecules that were used in its construction. The molecule itself is called the descendant.
Construction information, similar to the construction information presented in the Text Pane of Molecule Display Window, can be generated either for selected molecules only, or for selected molecules and all of their parent molecules. The latter option allows you to see the line of inheritance between the molecules.

Select **File > Reports > Relations** to generate a Relation Report.

**Step 1**

Step 1 in a Relational report takes place exactly as explained in Step 1 on page 376. Click **Next** to proceed with a new report or **Finish** if you are using a Relation template.

**Step 2**

In Step 2 of a Relational Report, you can select the relationship information to be included in this report. *For an example, see Figure 26.7.* Click **Next** or **Finish** in this window to proceed to the Preview window, similar to Figure 26.4. At any time, click **Back** to return to a previous screen.

![Figure 26.7 Selecting the relationship information to be included in the report](image)

**Report Templates**

To save any report that you generate as a template, click the **Save Template** button in the Report Preview window (Figure 26.4). In the Save Report Template dialog box that opens, name the new template. Click **OK** to save.

To use a saved report template to generate a new report, in an open Molecule Display window, select **File > Reports > <report type>**. In the Select Report Type dialog box that opens, select **Report based on existing template**. Select the template you wish to use and click **Next** to proceed. Then follow the steps beginning on page 376 to generate the report.

Report templates are stored separately for DNA/RNA molecules and proteins. To rename or delete a template, select **File > Reports > Templates**. Use the drop-down menu at the top of
the Templates window (Figure 26.8) to select the template type. Select the appropriate template and click **Delete** to remove it or **Rename** to rename the selected template.

![Figure 26.8 Templates window](image)
CHAPTER 27

SHARED DATABASES

Introduction

Special databases, repositories of DNA/RNA or protein molecules, enzymes, oligonucleotides, and gel markers, can be created and shared between several Vector NTI users on a network.

Shared databases are not a replacement for local databases; each Vector NTI application still must have its own local database. The local database is used in all operations: construction, design, creation of viewers etc. The local database is also the place for storing private and temporary data. The main purpose of the shared database is to store common data.

The only operations you can perform on shared databases are:

1. Copying data to/from the local database.
2. Various database management operations such as creating and deleting subsets, etc.
3. Database searches. For instance, you may copy some of your molecules and enzymes from your local database into the shared database. In order to use them in the design process, your colleague must copy them to his/her local database first.

In addition to biological data, each database contains information about its creator and registered users. Only the creator and registered users can have access to database data. The database creator can also change database properties, remove registered users, and define the password required to become a register user.

Note: Vector NTI shared databases use portable data format and file naming conventions to ensure that both Macintosh and Windows users of Vector NTI can access common databases. Shared databases can be located on a wide range of file servers—Vector NTI can work using not only services native to each system, (Microsoft Network or AppleTalk) but also various Unix (NFS or Samba) and NetWare services.

Creating a New Shared Database

Shared database capability is a purchased addition to Vector NTI Advance. When shared database capability is purchased, you are issued a Vector NTI shared database license that allows
you to create shared databases. A shared database license is a special type of static license that allows you to create any number of shared databases but it also limits the number of users for each database you create. You do not need a shared database license to access shared databases, but you should have a Vector NTI application version 3.1 or later.

**Note:** Shared databases can be accessed but cannot be created from Vector NTI using a license that is shared through a network server (Dynamic License).

Before creating a shared database, you should find a place on your network (a computer, storage volume and directory or folder) where the database will be located and accessible from the computers of all potential users of the database. Users should be able to connect to this location or mount the corresponding network drive and have read/write/delete permissions for all files in a database directory and its subdirectories. Good places for shared databases are file servers that can be "visible" as storage devices to other computers in a network. Vector NTI requires that a separate empty directory should be created for each shared database.

When the network directory for a new shared database is arranged, click the **Local/Shared Data Exchange** button ( ) or choose Database > Local/Shared Data Exchange. This opens the Connect To Shared Database dialog box (Figure 27.1) where you can perform a number of database management operations: create a new database, find the location of an existing database, and edit database properties and user information.

### Connecting to a Shared Database

In the Connect to Shared Database dialog box, click the **Create New Database** button.

In the Create dbdef file dialog box that opens, enter a name and select the destination for the shared database definition file (dbdef). Click **Save**.
In the Database Properties and Users dialog box that opens (Figure 27.2), you can provide additional information for the database users:

![Database Properties and Users dialog box](image)

- **Database Name**: enter the name of the new database. This name will be displayed in all dialog boxes and messages related to the database.

- **Password Required To Register At Login**: enter in the text box a registration password that will be required to become a registered user of this database. If you decided that the password is not necessary you can leave this field blank. If the password is not blank, Vector NTI will ask you to confirm the password before you leave the dialog box.

- **User Information**: enter appropriate user information

When you are finished press the **OK** button.

Vector NTI writes your data into the database definition file and creates the files and directories required to store the database data. The database you just created has no registered users. To be registered, a user should login to the database and enter the registration password you specified in the Password Required to Register at Login box. This procedure is described in details below.

**Note**: If anything goes wrong at this stage, check the network connection and make sure that you have “write permissions” for the chosen directory. Then, delete all files in the directory and repeat the last step.

In the Connect To Shared Database dialog box (Figure 27.1) you will see the name and location of the database you just created.

**Finding the Location of an Existing Shared Database**

If you are going to use an existing shared database you should specify the location of the database. First, make sure that you have a working network connection to this location or that the corresponding network drive is mounted and you have read/write/delete permissions for all files in the database directory and its subdirectories.

**Note**: On some systems you can mount network drives from the standard File dialog boxes. In this case, no separate step is required.

In the Connect To Shared Database dialog box (Figure 27.1), click the **Find Database** button. In the Find dbdef file dialog box, locate the shared database definition file (dbdef). Find the database directory and click the **Open** button. Vector NTI tries to read the database definition file; if
it succeeds, you will see the name and location of the database in the Connect To Shared Database dialog box.

Editing Database Properties and User Accounts

When the location of a shared database you created is specified, you can edit the database name and registration password. The creator of the database can also remove registered users. To edit database properties and user account information, in the Connect to Shared Database dialog box (Figure 27.1), click the **Edit Database Properties and User Accounts** button, opening the Database Properties And Users dialog box where you can modify the corresponding database information.

Saving the Connections in a Connection Profile

When the location of a shared database is specified, you can enter login information that will be displayed to other users of a shared database. This is a good way to communicate your personal information such as name and phone number. When you are logged into a shared database, Vector NTI displays this information in addition to the login time and the information about your computer to other database users attempting to log in. In order to return to this configuration later, you may save both database location and login information in a connection profile.

To save the connections in a connection profile, in the Connect to Shared Database dialog box (Figure 27.1), click the **Save Connection As** button and enter the name of the connection in the Save Connection As dialog box. Vector NTI stores the connection profile in the preference file; each time you open the Connect To Shared Database dialog box you see the stored connections list near the top of the dialog box. You can select stored connections from this list. To delete the connection from the profile, click the **Remove Connection** button.

Connecting to a Shared Database

When the location of a shared database is specified, to connect to the database, click the **Connect** button. Vector NTI reads the information about the database creator and registered users. If you are not a creator or registered user of the database and the database limit of registered users is not reached, Vector NTI displays a dialog box allowing you to register.

To register, choose **Yes**. If the user registration is protected by registration password, you will be asked to enter the registration password in the Login Registration dialog box. If the password is correct, Vector NTI adds you to the list of registered users of the shared database.

If you are the database creator or registered user, you have full access rights to the database data. However, if you are not going to write to the shared database this time, you can check **Read Only** box. This guarantees that the shared database contents will not be accidentally modified.

If you are not the database creator or registered user, you can choose to connect to the database in a browse-only mode to look at the database contents.

**Note:** If Vector NTI cannot access the database at this or the subsequent stages, check the network connection and make sure that you have write permissions for the files in database directory and its subdirectories.

Managing a Shared Database

When your connection to the selected shared database is established, Vector NTI opens a separate Database Explorer window for the shared database (Figure 27.3). This window is a close functional equivalent of the local Database Explorer window. In this shared Database Explorer, you can search the database, manage subsets, rename and delete objects etc. However, the concept of Vector NTI's shared databases prevents you from manually creating new objects in
shared databases, opening Display Windows for objects stored in shared databases and editing shared database objects in place using object editor dialogs.

![Figure 27.3 Database Explorer window for the shared database](image)

To modify a shared database object, you must copy it to your local database, modify it there and put the modified copy back into the shared database as described in the next section.

**Note:** If the original object was submitted by another user, instead of replacing it, Vector NTI creates a separate object. This new object has the same name but a different “Submitted by” field, containing the Author Name you used to identify yourself in the Author Information dialog box.

### Exchanging Data with a Shared Database

Exchanging data between your local and shared databases is simple and intuitive. Open two Explorer windows (one for local and another for shared database), then copy data between them using drag-drop or copy/paste operations described in the Copying Objects section. The simplest way to start data exchange is to select **Database > Local/Shared Data Exchange** or click the **Local/Shared Data Exchange** button in the Database Explorer window. This command allows you to choose a shared database, connects to it (as described in the previous sections), opens the shared database Explorer window and the local Explorer window (if needed) and then tiles two Explorer windows on your screen.

You can place your objects into a shared database without worrying about possible name conflicts with objects placed there by your colleagues. Vector NTI automatically stores the name you used to identify yourself in the Author Information with the objects you put into a shared database. This name is stored in the shared database in the “Submitted by” field of an object (it can be different from both the Author and Original Author fields of the object). In addition, Vector NTI automatically creates and maintains special subsets containing objects that were submitted by the same users. The name alone is no longer a unique way to identify an object in the shared database; two or more different objects can have the same name provided that they came from different sources.

### Disconnecting from a Shared Database

When you are finished exchanging data in the Shared Database Exchange dialog, click the **Close** button. You will be disconnected from the shared database and the database can be used by another user.
Introduction

The simultaneous alignment of many nucleotide or amino acid sequences is an essential tool in molecular biology. Multiple alignments are used to find diagnostic patterns, to characterize protein families, and to detect or demonstrate a similarity between new sequences and existing families of sequences. They are also useful for predicting secondary and tertiary structures of new sequences, for suggesting oligonucleotide primers for PCR and serving as an essential prelude to molecular evolutionary analysis.

AlignX, the multiple sequence alignment application of Vector NTI Advance, is a comprehensive program designed to conduct and manage multiple sequence alignment projects. It uses a modified Clustal W algorithm (see page 402) and incorporates the following features:

- Profile alignment
- Guide tree construction, displayed in graphical representation
- Use of residue substitution matrices
- Secondary structure consideration
- Multicolored alignment presentation
- Automatic consensus calculation
- Full alignment editing capabilities
- Dot Matrix comparison of any two sequences

Launching Align X

To launch AlignX from the Start menu, select Start > Programs > Invitrogen > Vector NTI Suite 10 > AlignX. From Vector NTI Database Explorer, select Align > AlignX—Open New Alignment Window or select a group of molecules then choose Align > AlignX—Align Selected Molecules. You can also drag the selected molecules from the Database Explorer into the AlignX workspace. From other Vector NTI applications, select the molecules to be included in the alignment and choose Align > AlignX – Align Selected Molecules.
Features of the AlignX Workspace

When AlignX is first opened, the display window is empty. As molecules are added to an AlignX project, they are listed in the Text Pane. After aligning a group of molecules selected in the Text Pane, alignment results then appear in the Graphics Pane and the Alignment Pane.

Because it will be much easier for you to understand the AlignX features in a complete display window, open a demo project into the workspace. Select Project > Open. Find the Vector NTI Suite folder. (The default location is Program Files\Vector NTI Suite 10\Demo Projects.) Open the Protein.apr file, a sample alignment of 25 molecules.

The AlignX interface (Figure 28.1) consists of a menu bar, three toolbars, an AlignX Display window, with shortcut menus specific for each pane. A Dot Matrix window can be opened from the AlignX display window when the Dot Matrix tool is launched.

![Figure 28.1 AlignX interface](image)

AlignX has three toolbars: the Main Toolbar, the Pane Selection Toolbar and a Pane Toolbar. The Main Toolbar contains tools for major program functions, such as opening and saving projects, adding molecules to the project, exporting molecules' data in various formats, and creating alignments.

As you are introduced to various operations in this chapter, the toolbar buttons you will use are displayed appropriately. Most toolbar buttons have associated tool tips that explain the button function; hover the cursor over the button to display the tool tip for that button. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.

The AlignX Display window is divided into four panes separated by split bars: a Text Pane, a Guide Tree Pane, a Graphics Pane, and an Alignment Pane. To apply a command from the main menu in a given pane (i.e. Print, Print Preview, Camera), that pane must be active. The active pane can be toggled with the Switch Panes buttons (,) in the Pane Selection Toolbar or by clicking anywhere in the pane you want to activate.
Split bars allow re-sizing of the four panes in the AlignX Display window. Scroll bars appear when the contents of a pane exceed the viewing area, facilitating easier viewing of pane contents.

Text Pane

The Text Pane contains a thorough description of each molecule included in the alignment project. The information is organized in folders, named by molecule. Each top level (molecule) folder contains the following subfolders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description</td>
<td>Molecule type, form (for DNA only), length, etc.</td>
</tr>
<tr>
<td>Proprietary fields</td>
<td>User defined fields (for molecules which came from Vector NTI)</td>
</tr>
<tr>
<td>Standard fields</td>
<td>GenBank/SWISS-PROT-like fields: keywords, division, original accession numbers, etc.</td>
</tr>
<tr>
<td>Comments</td>
<td>Arbitrary text of any length associated with the molecule</td>
</tr>
<tr>
<td>References</td>
<td>Bibliographic references (in GenBank/SWISS-PROT format)</td>
</tr>
<tr>
<td>Analysis (Protein Molecules only)</td>
<td>Results of common physiochemical analyses</td>
</tr>
<tr>
<td>Feature Map</td>
<td>List of molecule features and their positions on the molecule</td>
</tr>
</tbody>
</table>

*Table 28.1  AlignX Text Pane folders*

Folders and sub-folders can be opened with a double-click or by clicking on the + to the left of the folder name.

Refer to Chapter 3 for more information regarding the following Text Pane manipulations:

- Expanding and collapsing folders
- Shortcut menus
- Copying pane contents to the clipboard

**Finding Features on Graphical and Alignment Panes**

Features listed in the Feature Map subfolder in the Text Pane can be highlighted on molecule graphics and/or sequences in the Graphics and Alignment Panes (only for those molecules that are part of the current assembly). To visualize the positions of a signal in these panes, select the feature in the Text Pane folder, then click the Find button ( ) on the Text Pane toolbar or select Edit > Find Feature on the menu bar. *The Find command is disabled if selected line is not a signal with non-zero boundaries.*

**Marking Features in the Alignment Pane**

Features listed in the Feature Map subfolder in the Text Pane can be marked graphically in the Alignment Pane (for those molecules that are part of the current alignment).

To mark a particular feature in the Alignment Pane, right-click on the feature of interest in the Text Pane folder and choose Show Feature on Alignment Pane from the shortcut menu. A
graphic depiction of the feature displays below the molecule sequence where the feature occurs (Figure 28.2):

![Figure 28.2 Feature displayed graphically in the Alignment Pane](image)

To mark all the features listed in the Feature Map subfolder for a particular molecule, right-click on the molecule of interest in the Text Pane and choose **Show all Features for <selected molecule name>** from the shortcut menu. All the features listed in the Feature Map subfolder of the molecule display graphically in the Alignment Pane under the molecule sequence.

To hide a feature or all features for a molecule displayed in the Alignment Pane, right-click on the feature or molecule, respectively, in the Text Pane, as described above, and choose the corresponding Hide command (**Hide Feature on Alignment Pane** or **Hide all Features for <selected molecule name>**, respectively) from the shortcut menu.

**Other Feature-Related Commands**

Various other feature-related commands are available on shortcut menus launched from the Alignment Pane. For more information on these commands, see Displaying Molecule Features in the Alignment Pane, page 396.

**Guide Tree Pane**

Phylogenetic analysis is the means of studying presumed evolutionary relationships. An inferred evolutionary history is displayed in a treelike diagram suggesting the inherited relationships between the molecules.

A guide tree, which resembles a Phylogenetic tree, in Vector NTI Advance is built using the Neighbor Joining method (NJ) of Saitou and Nei. The NJ method works on a matrix of distances between all pairs of sequence to be analyzed. These distances are related to the degree of divergence between the sequences. The guide tree is calculated after the sequences are aligned. AlignX displays the calculated distance values in parenthesis following the molecule name in the Guide Tree Pane.

If there are more than 3 sequences in an alignment, a guide tree is shown in the Guide Tree Pane. To activate this pane, click the **Guide Tree** button on the Pane Selection Toolbar. Scrollbars in the Guide Tree Pane enable you to view the entire tree.

**Exporting a Guide Tree**

To export the guide tree into a .ph file for use with other applications, click the **Export Tree** button on the Guide Tree Pane Toolbar or choose **View > Export Guide Tree** on the menu bar or select **Export Guide Tree** from the shortcut menu. Select or enter the file name to export to in the File Save dialog and click **Save**.

**Interconnection with Other Panes**

Any molecule selected in the Guide Tree Pane is selected concurrently in the Alignment Pane.
Graphics Pane

To activate the Graphics Pane, click on the **Graphic Pane** button ( ) on the Pane Selection Toolbar. The graphical representation of an alignment contains three graphs in the Graphics Pane. Graphs are divided by horizontal split bars that can be moved to change the height of the display area.

**Descriptions of the Graphs**

- **Alignment Similarity** (upper graph): Specific values (in a 0-1 range) are assigned to each residue at a given alignment position in each aligned sequence, depending on whether the residue is identical, similar or weakly similar to the corresponding residue of the consensus sequence. The values (1 (identical), 0.5 (similar), and 0.2 (weakly similar) for each residue at a given position are totaled; the sum is divided by the number of the sequences in the alignment, “normalizing” the resulting value.

- **Absolute Complexity** profile of the Multiple Alignment (center graph): This graph is calculated as a sum of all pairwise residue substitution scores at a given alignment position divided by the number of pairs in the alignment. The scores are taken from the residue substitution matrix used for alignment calculation.

- **Absolute Complexity** profile of the Pairwise Alignment for the molecule selected in the Guide Tree/Alignment Pane relative to the consensus sequence (lower graph): This graph is calculated in a manner identical to the one in the second graph. Selection of a molecule in any pane results in the calculation of this graph. This graph is the only one that changes according to which molecule is selected.

All graphs display the values averaged in a window of a specific length (defined by “window” parameter) that slides along the alignment.

You can add additional analyses to the graphics pane by activating the Graphics Pane and choosing **View > List of Analyses** from the drop down menu to open the Analysis List setup dialog box (Figure 28.3):

*Figure 28.3 DNA/RNA Analysis List Setup dialog box*

The dialog box lists all of the analyses available for DNA or Protein, depending on the type of aligned molecules. Currently displayed analyses are listed in the Selected Analyses box. To add an analysis, select it in the Available Analyses list box and click the **Add Analysis** button ( ).
to move it to the Selected Analyses list. Alternately, you can move an analysis to the lower list by double clicking on it.

The plot display order is determined by the order the analyses are listed in the Analysis List setup box. To modify the order, select an analysis and move it up by clicking the **Move Up** button ( ). Click **OK** to close the Analysis list setup dialog box and generate the new analyses.

**Adding Custom Protein Analyses**

Specific values are used in the calculation of the various protein analyses. These values differ depending on the particular analysis and are stored in small text files having a .pa2 extension. All .pa2 analysis files are stored in the Analyses subdirectory in the Vector NTI Suite directory. The name of the analysis corresponds with the name of the .pa2 file.

For example, the Bulkiness analysis is characterized in the Bulkiness.pa2 file as follows:

```
[general]
author=Zimmerman J.M., Eliezer N., Simha R.
comment=Last modified 15/Mar/1996 by ELG
unique=s30
[data]
```

The [general] section describes the attributes of the analysis and the [data] section contains the list of values used for calculating the analysis. Values are listed according to the alphabetical order of the acid coefficients.

You can create your own custom protein analysis by modifying the values in one of the existing .pa2 files using a plain text editor, such as Notepad. Once modified, the new .pa2 file should be saved in the Analyses subdirectory under a new name that corresponds to the new analysis type. If AlignX is open at the time the new .pa2 file is created, it must be closed and reopened before the new analysis is recognized.

**Plot Setup**

To modify a particular graph in the Graphics Pane, right click on the graph you want to change and choose **Plot Setup** from the shortcut menu.

![Figure 28.4 Analysis tab of Plot Setup dialog box](image)
On the **Analysis** tab of the Plot Setup dialog box (Figure 28.4), you can set the Window size, that is the number of sequential residues used to calculate the average for each graph point. The larger the number entered (Window size), the smoother the graph curve.

On the **Datagram** tab of the Plot Setup dialog box (Figure 28.5), you can modify the plot color. Both the size of the window and plot color you select are specific for the selected graph in the current project. You can also change the graph style and choose to show or hide the horizontal axis for that graph here. These settings are stored within the project file and applied to graph next time you load the project.

**Zooming Graphs In and Out**

Enlarging or reducing regions of alignment graphs is possible using the Zoom buttons. For descriptions of their use, refer to Chapter 3.

**Managing the Scales of the Graphs**

Each graph has its own vertical and horizontal scales. Vertical scales are similar for each graph and show the maximum and minimum values for the graph. On the horizontal scale, either numerical positions in the sequence or residues can be shown, depending on the degree of zooming in the graphics pane. If you would like to see the residue designations, zoom in on the graphics pane. The vertical and horizontal scales can be further managed as follows:

- As mentioned above, the horizontal axis can be shown or hidden for a particular graph by checking or unchecking the Show individual axis box on the **Datagram** tab of the Plot Setup dialog box.

- The **Show Vertical Axis** command is used to show or hide the vertical axis for all the graphs. This command can be accessed from a button on the tool bar or the shortcut menu launched by right clicking on the Graphics pane.

Graph legends can be removed from the pane by unselecting the **Show Legends** button from the Pane Toolbar or by unchecking the **Show Legend** command from the shortcut menu or the View menu. The **Place Legends Below Plots** command is used to toggle the graph legends display from the right side of the graphs to below the graphs. The **Place Legends Below** Plots command can be accessed either from the shortcut menu or from a button on the toolbar.

If no items are checked, the graphs are shown without annotation.
Interconnection with Alignment Pane

Selections within the Graphics Pane are concurrently selected in the Alignment Pane. In general, regions can be selected using the click + drag technique. Detailed selection techniques are discussed in Chapter 3.

Alignment Pane

To activate the Alignment Pane, click on the **Alignment Pane** button ( ) on the Pane Selection Toolbar.

The Alignment Pane displays aligned sequences and the resulting consensus sequence. Consensus residues are those that appear most commonly at a particular site. The Alignment Pane has its own vertical and horizontal scrollbars. The last row in the pane consists of the alignment consensus.

Vertical scrolling affects both the aligned sequences and their names, but the consensus sequence is not included in the scroll region: it remains visible as the bottom sequence. The sequence names remain stationary at the left of the Alignment Pane no matter how the sequences are repositioned horizontally. If the list of molecules listed in the Alignment Pane is longer than the size of the pane, a vertical scroll bar appears as well. Vertical scrolling affects both sequences and their names but the consensus sequence remains visible at the bottom of the pane.

AlignX shows similarity and identity levels for the generated alignment in the Alignment Pane. The Status bar displays the percent of similar and identical residues for entire alignment or for a selected region on the alignment. When the mouse cursor is pointed to any particular column within alignment, a popup label shows the percent of similar and identical residues within an alignment column. The popup label option is disabled if only two sequences are aligned.

Interconnection with Other Panes

When a molecule is selected in the Alignment Pane, it is selected in the Guide Tree Pane as well. A guide tree is not calculated when there are three or less sequences in the alignment.

A selected region within the Alignment Pane is also selected concurrently in the Graphics Pane.

Displaying Molecule Features in the Alignment Pane

Features listed in the Feature Map subfolder in the Text Pane can be marked graphically in the Alignment Pane (for those molecules that are part of the current alignment) using commands executed from the Text Pane (see Text Pane, page 391). Once features have been marked in this way, a number of feature related commands are enabled on the Alignment Pane right-click shortcut menus as follows:

Graphical Feature Shortcut Menu Options

The following commands are available on the shortcut menu launched by right-clicking on a graphical feature shown in the Alignment Pane.
<feature name (position)> – launches the Feature Type Color Scheme dialog box where you can change the color scheme for the selected feature type or any other feature type (Figure 28.6):

![Feature Type Color Scheme dialog box](image)

**Figure 28.6 Feature Type Color Scheme dialog box**

Select the feature type you want to change the color of on the left side of the dialog box and choose the new color from the color drop-down box for that feature on the right side of the dialog box.

- **Hide <feature name>** – removes the feature graphic from the alignment display.
- **Copy to Consensus <feature name>** – copies the feature graphic to the consensus sequence display.
- **Move to New Line** – moves to a new line a graphical feature that occurs on the same line as another graphical feature.

The following commands are available on the shortcut menu launched by right-clicking on a graphical feature shown in the Alignment Pane for the Consensus sequence ONLY:

- **Delete from the Consensus <feature name>** – deletes the selected graphical feature from the Consensus sequence display.
- **Edit <feature name>** – allows you to change the name of the selected Consensus sequence graphical feature.

**Alignment Pane Shortcut Menu Options**

The following commands are available on the shortcut menu launched by right-clicking anywhere in the Alignment Pane:

- **Delete Selected Features** – deletes the currently selected graphical feature from the alignment display.
- **Feature Type Color Scheme** – launches the Feature Type Color Scheme dialog box (Figure 28.6) where you can change the color display for feature types.
- **Create Feature** – allows you to create a graphical feature for the consensus sequence. The portion of the consensus sequence you want to mark must be highlighted before the Create Feature command is enabled.
Finding Sequences in the Alignment Pane

You can perform sequence searches on molecules in the Alignment Pane using the AlignX Find Sequence feature. To open the Find Sequence dialog box (Figure 28.7), right-click on the Alignment Pane and choose Find Sequence from the shortcut menu.

![Figure 28.7 Find Sequence dialog box](image)

Select the molecule you want to search by highlighting its name on the left side of the Alignment Pane. Type the sequence you want to search for in the Find Sequence text box. Set the search parameters to achieve the desired search direction and looping status:

- **Forward** – Searches forward from the cursor or highlighted area of the selected molecule, stopping at each point where the entered sequence string is found. Once the selected sequence has been searched to its end, the search continues with the next molecule down the list. This process continues until the last molecule in the list, the Consensus, has been searched.

- **Backward** – Searches backward from the cursor or highlighted area of the selected molecule, stopping at each point where the entered sequence string is found. Once the selected sequence has been searched to its beginning, the search continues with the next molecule up the list. This process continues until the first molecule in the list has been searched.

- **Loop Search** – When a selection is made on a sequence in the Alignment Pane, the search begins at the selection. The Loop Search option loops the search back to the beginning of the selection.

Once you have set the search conditions appropriately, click the **Find** button to begin the search.

Copying the Reverse/Reverse Complement of a Molecule

Although you can copy to the Clipboard the reverse sequence (Protein) or reverse complement (DNA/RNA) of a molecule (or selected portion of a molecule) in Vector NTI and ContigExpress, this feature is not available in AlignX.

**Note:** You can use the Reverse Complement feature employed in Vector NTI (Chapter 19) and ContigExpress (Chapter 30) with a molecule, and then subsequently add that molecule to an open AlignX project.

From AlignX, save the desired molecule to the database by selecting **File > Save As** (possible file extensions are ".txt", ".gb", ".gp"). Then open the molecule in Vector NTI or ContigExpress and follow the directions for making the reverse complement of the molecule as described in Making the Reverse Complement of a DNA Molecule, page 244 or Making a Reverse Complement, page 454.) Save the reverse copy/reverse complement of the molecule.

Once the reverse copy/reverse complement is saved, add it to an active AlignX project by selecting **Project > Add Files** (described on page 401 in this chapter). Locate the newly created reverse copy/reverse complement file and click **Open**, adding the reverse copy/reverse complement to the project.
**Changing Molecule Order in Alignment Pane**

There is one way to reorder the molecules in the Alignment Pane. Changes are retained in the AlignX project upon saving.

Moving Individual Molecules

To move individual molecules, click the molecule name in the Alignment Pane list and drag it to the desired position in the list.

**Alignment Display Setup**

To change a default set of alignment display parameters, click the **Alignment Display Setup** button on the Alignment Pane toolbar or choose **View > Display Setup** from the menu bar or **Display Setup** from the shortcut menu. The Alignment Display Setup dialog box opens with three alignment property tabs: Consensus Calculation, Similarity Tables, and Color Setup.

Consensus Calculation Tab

A consensus sequence is a theoretical representative nucleotide sequence in which each nucleotide represents either the residue seen most frequently at that same site in aligned sequences, or is selected by other criteria. The Consensus Calculation tab specifies how the consensus sequence, displayed as the bottom sequence in the Alignment Pane, is calculated in AlignX.

![Figure 28.8 Consensus Calculation tab of Alignment Display Setup dialog box](image)

On the Consensus Calculation tab (Figure 28.8), you can select among the following parameters:

- Consider only identical residues...
- Ignore gaps in consensus calculation
- Specify a sequence to use as the consensus sequence. In this case, select one of the sequences listed (those used in the alignment).
- Residue fraction for consensus

If the **Use sequence as consensus** checkbox is checked, the Consensus will not be calculated but the sequence selected in the listbox below the checkbox will be used as the consensus.
Similarity Tables Tab

The Similarity Table (Figure 28.9) allows you to review or to define residue similarity information and values necessary for calculation of the alignment quality profile. Each residue pair can be defined as being strongly or weakly similar to each other. The table is editable only for amino acid residues.

![Similarity Tables tab of Alignment Display Setup dialog box](image)

Color Setup Tab

Residues in an alignment are colored according to the following scheme:

<table>
<thead>
<tr>
<th>Color</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>black on window</td>
<td>non-similar residues</td>
</tr>
<tr>
<td>default color</td>
<td></td>
</tr>
<tr>
<td>blue on cyan</td>
<td>consensus residue derived from a block of similar residues at a given position</td>
</tr>
<tr>
<td>black on green</td>
<td>consensus residue derived from the occurrence of greater than 50% of a single residue at a given position</td>
</tr>
<tr>
<td>red on yellow</td>
<td>consensus residue derived from a completely conserved residue at a given position</td>
</tr>
<tr>
<td>green on window</td>
<td>residue weakly similar to consensus residue at a given position</td>
</tr>
<tr>
<td>default color</td>
<td></td>
</tr>
</tbody>
</table>

![Table 28.2 Alignment default color scheme descriptions](image)
To modify alignment display colors, in the Alignment Display Setup dialog box, select the Color Setup tab (Figure 28.10).

![Figure 28.10 Color Setup tab of Alignment Display Setup dialog box](image)

Change the colors for the specific project using the drop-down menus, observing the modifications in the preview box. Click OK to save color changes in the alignment project. They will be applied to the alignment next time you open the project.

After any changes are saved, consensus and colors for residues in the Alignment Pane are recalculated as well as values for all the three graphs in the Graphics Pane, and both panes are redrawn.

**AlignX Projects**

**Creating a New Project**

When AlignX is launched, a new, empty alignment workspace is created. To create a new alignment project while in the AlignX display window, select Project > New on the menu bar. If you have any unsaved changes in an existing, and currently open project, AlignX suggests changes should be saved.

In the dialog box that opens, select Yes to save changes, No to disregard changes or Cancel to discard changes and continue work on the opened project. In the first two cases, the current project will be closed and the workspace emptied.

**Opening An Existing Project**

To open an existing project, choose Project > Open on the menu bar or click the Open Project button ( ) on the Main Toolbar. In the Open dialog box, select the desired file and click Open. By default, alignment projects have an “.apr” file extension. If the file is an alignment project, information about molecules, properties and alignments (if any) appears in appropriate panes.

Another way to open an existing project is to drag and drop the project file on the AlignX workspace.

**Adding Molecules to the Project**

Molecules for a new (or existing) project can be added to AlignX in these ways:

- In the Vector NTI Database Explorer, in either a DNA/RNA subset or a protein subset, select a list of molecules, and drag them into the Text Pane of the empty AlignX workspace. Molecules can be dropped in the AlignX workspace to be added to the project if they are of an appropriate file format and molecule type.
Select **Project > Add files** on the AlignX menu bar or click the **Add Files** button ( ) on the Main Toolbar. Choose the appropriate file of molecules from the Add files to project dialog box, and click **Open**.

AlignX reads various file formats: GenBank and EMBL for DNA, GenPept and SWISS-PROT for proteins, FASTA and Vector NTI archives for both DNA and proteins. The first added molecule determines the molecule type. If you try to add molecules of a different type later, a warning message appears and the inappropriate molecule is not added. If a molecule’s type correlates with the current project, the molecule(s) are added to the project and appear in the Text Pane.

If the program fails to define a file’s format, the file can be imported as a raw sequence. If there are no molecules in the project, the molecule’s type must be selected before it is possible to import the file (Figure 28.11). If the project is not empty, the molecule type is already defined and the raw sequence can be imported only if its file type matches the current project.

Figure 28.11 Importing a raw sequence

Imported raw sequences appear in the Text Pane with the names such as NONAME, and NONAME#2. Select **Rename Molecule** from the shortcut menu to rename the molecule.

Even after molecules are brought in the AlignX workspace, three of the panes are empty until the alignment is created.

### Alignment Import in MSF Format

In addition to its own projects, AlignX can open alignments of the MSF format. To open an MSF project, choose **Project > Import MSF Format** from the menu bar. In the Import MSF Format dialog box, select the MSF project to be opened. By default, MSF projects have an .msf extension, but are not limited to it. AlignX tries to recognize the MSF format itself, not simply by its file extension. You can also open an MSF project by dragging the file and dropping it within the AlignX window. An MSF project imported into a new AlignX project is initially named **untitled**. Select **Rename Molecule** from the shortcut menu to rename the molecule.

The information contained within an MSF project is molecule type, name, length and sequence. For this reason, Text Pane folders of MSF-format molecules contain no information about creation and modification date, author, original access number, references, and feature tables. Work is done with an MSF project just like any other AlignX project.

### Initiating an Alignment

There are two ways to initiate alignments. In the Text Pane, select the sequences to be included in the alignment, then proceed with either a multiple alignment, or a profile alignment.

### Multiple Alignment

In a multiple sequence alignment in AlignX, all sequences included in the alignment are “equally important.” AlignX uses a modified Clustal W algorithm in which pairwise alignments are performed between sequences and the alignment tree proceeds using the particular scoring matrix designated by the user in Alignment Setup (see page 407), whereas in the traditional Clustal W,
the algorithm chooses the scoring matrix or series of matrices as the alignment progresses. (See page 661 for alignment algorithm references.)

To initiate a multiple alignment, click the Align button on the Main Toolbar or select Align > Align Selected Sequences from the menu bar. This function is enabled only if more than one molecule is selected in the Text Pane.

Profile Alignment

A profile alignment in AlignX is based on aligning all selected sequences against a reference sequence. This is based on the Clustal W algorithm. To initiate a profile alignment, click the Align Using Profile button on the Main Toolbar or choose Align > Align Selected Using Profile on the menu bar. In the Select Profile dialog box (Figure 28.12) that opens, select the molecule to be used as the first profile.

![Select Profile dialog box](image)

Figure 28.12 Select Profile dialog box

This function is available only if more than 1 molecule is selected in the text view.

Adding a Molecule to an Alignment

To add a molecule to the alignment, select the molecule in the Text Pane or drag it in from Database Explorer. Click the Add to Alignment button or choose Align > Add Selected To Alignment. A profile alignment using an existing alignment as the first profile will be performed. This function is available when an alignment already exists and at least one molecule that does not belong to alignment is selected in the Text Pane.

Realigning a Selected Area of an Alignment

To realign the selected portion of an alignment, in the Alignment Pane, select the region of the alignment you want to realign by clicking and dragging over the area of interest. Right-click in the Alignment Pane and choose Realign Selected Area from the shortcut menu. AlignX realigns the selected area; the unselected regions of the alignment retain their original alignment.

Removing a Molecule from the Project

To delete a selected molecule from the project, click the Delete from Project button from the Text Pane toolbar, choose Edit > Delete <molecule name> From Project on the menu bar or Delete <molecule name> From Project from the shortcut menu. If you confirm this operation, the specified molecule will be deleted from the project. If the molecule is removed from the alignment, the guide tree, consensus and graphs are recalculated and refreshed.

Setting Alignment Parameters

The first molecule added to an alignment defines the molecule type of the alignment (DNA or protein). When initiated, a default set of alignment options is created for the defined molecule.
type. To change these options, click the **Alignment Setup** button ( ) on the Main Toolbar or choose **Align > Alignment Setup**. The Alignment Parameters dialog box opens with five possible property tabs: Pairwise Alignment, Multiple Alignment, Protein Profile Alignment, Tree Construction Options and Score Matrix (Figure 28.13).

**Pairwise Alignment**

The method for calculating pairwise distances is defined on the Pairwise Alignment tab.

![Figure 28.13 Alignment Setup dialog box, with Pairwise Alignment tab displayed](image)

Pairwise alignment parameters control the speed/sensitivity of the initial alignments:

- Fast (approximate) method
- Slow (more accurate) method uses two gap penalties (for opening or extending gaps) and a full amino acid weight matrix.

By default, the slower method is used. There are two groups of parameters enabled on this pairwise tab depending upon which method is chosen.

**Note:** Default settings are in parentheses: first number = DNA; second number = protein

<table>
<thead>
<tr>
<th><strong>Pairwise Alignment Parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Slow Options</strong></td>
</tr>
<tr>
<td>Gap opening penalty (15/10)</td>
</tr>
<tr>
<td>Gap extension penalty (6.66/0.1)</td>
</tr>
<tr>
<td><strong>Fast Options</strong></td>
</tr>
<tr>
<td>Number of K-tuple matches (2/1)</td>
</tr>
<tr>
<td>Number of best diagonals (4/5)</td>
</tr>
<tr>
<td>Window size (4/5)</td>
</tr>
<tr>
<td>Gap penalty (5/3)</td>
</tr>
</tbody>
</table>

*Table 28.3 Pairwise Alignment parameters*
Multiple Alignment

These parameters control the final multiple alignment (Figure 28.14).

Figure 28.14 Multiple Alignment tab of Alignment Setup dialog box

Note: Default settings are in parentheses: first number = DNA; second number = protein

<table>
<thead>
<tr>
<th><strong>Multiple Alignment Parameters</strong></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gap opening penalty (15/10)</td>
<td>The penalty for the first residue in a gap</td>
</tr>
<tr>
<td>Gap extension penalty (6.66/0.05)</td>
<td>The penalty for additional residues in a gap</td>
</tr>
<tr>
<td>Gap separation penalty range (8)</td>
<td>Tries to decrease the distances between gaps</td>
</tr>
<tr>
<td>Use end gap separation penalty</td>
<td>Does not penalize for gaps introduced at the end of a sequence</td>
</tr>
<tr>
<td>% identity for alignment delay (40)</td>
<td>Delays the alignment of the most divergent sequences; that is, sequences that are less identical than the percentage shown will be introduced into the alignment later.</td>
</tr>
<tr>
<td>Use residue-specific gaps</td>
<td>Amino acid-specific penalties that increase or reduce penalties at each position in the aligned sequence</td>
</tr>
<tr>
<td>Use hydrophilic residue gap</td>
<td>Increases the chance of a gap within a region of hydrophilic residues (if yes, specify in the following text box)</td>
</tr>
<tr>
<td>Transition weighing</td>
<td>(For DNA only) Gives translations (A&lt;- &gt; G or C&lt;- &gt; T) a specific weight. (0) means the transition is treated as a mismatch while 1 gives the transition a match score (0)</td>
</tr>
</tbody>
</table>

Table 28.4 Multiple Alignment parameters
Protein Profile Alignment

This dialog box allows you to raise penalties for opening gaps in secondary structure (protein) regions (Figure 28.15).

![Protein Profile tab of Alignment Setup dialog box](image)

**Figure 28.15 Protein Profile tab of Alignment Setup dialog box**

**Note:** Default settings are in parentheses.

<table>
<thead>
<tr>
<th>Profile Alignment Parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Use secondary structure in first/second profile (yes or no)</td>
<td></td>
</tr>
<tr>
<td>Determines if the secondary structure information is to be used in the indicated profile</td>
<td></td>
</tr>
<tr>
<td>Gap penalty for helix core residue (4)</td>
<td></td>
</tr>
<tr>
<td>For helix core residues breaking</td>
<td></td>
</tr>
<tr>
<td>Gap penalty for strand core residue (4)</td>
<td></td>
</tr>
<tr>
<td>Gap penalty for structure termini (2)</td>
<td></td>
</tr>
<tr>
<td>For introducing gaps at the end of a secondary structure</td>
<td></td>
</tr>
<tr>
<td>Gap penalty for loop regions (1)</td>
<td></td>
</tr>
<tr>
<td>Raises the penalty for introducing gaps in a loop (By default this penalty is not raised.)</td>
<td></td>
</tr>
<tr>
<td>Number of residues inside/outside a helix/strand to be treated as terminal</td>
<td></td>
</tr>
<tr>
<td>Specifies the range of residues inside or outside a helix or strand beyond the structure</td>
<td></td>
</tr>
</tbody>
</table>

*Table 28.5 Profile Alignment parameters*

Guide Tree Construction Options

A guide tree is calculated from the alignment using the Neighbor Joining (NJ) method. On the Tree tab (Figure 28.16), two parameters may be changed for tree construction (neither of these are set initially):
Score Matrix

The Score Matrix tab (Figure 28.17) displays the matrix used to generate the current multiple alignment. From this tab, you can also open other matrices for review or for application to the alignment.

When a project is created, a residue substitution matrix from the corresponding template (DNA or protein) is used. If either template file is not found, the default matrix is used (swgapdnamt for DNA and blosum62mt2 for protein). To review other matrices for either DNA or proteins, clicking...
the Select Matrix button. From the Matrices file, select any other matrix to be loaded on the Score Matrix tab.

Editing a matrix takes place in the Matrix Editor, accessible from the program files where you installed Vector NTI Advance. For further details regarding customizing or editing matrices, refer to Chapter 39.

Editing the Alignment

To edit an alignment, click the Edit Alignment button ( ) on the Alignment Pane toolbar, or choose View > Edit Alignment from the menu bar or Edit Alignment from the shortcut menu. This opens the Edit Alignment dialog box containing rows of molecule names and sequences (Figure 28.18).

![Figure 28.18 Edit Alignment dialog box](image)

**Note:** You can resize the Edit Alignment dialog box by clicking and dragging on the lower right corner.

To edit the alignment, select the desired region in the sequence by LEFT-CLICK + DRAG through the sequence in the appropriate direction with the mouse or double-click with the left mouse button within the sequence, selecting the entire region between two gaps. The current selection, as well as its position in the sequence, is shown in the status bar of the dialog box.

Editing can only occur by shifting selected sequences into adjacent gaps. Therefore, the four arrow buttons allowing movement of the selected block are enabled only when there is a gap to the left or right of the selected region. To shift the block, in the Move Selected Block area, designate the number of positions to move in the text box and press one or more of the following arrow buttons:
- to the beginning of the adjacent gap interval in the left direction

- moves the number of positions designated to the left

- moves the number of positions designated to the right

- to the end of the adjacent gap interval in the right direction

Click the **OK** button to save any changes and close the Alignment Edit dialog box. Click the **Apply** button to save any changes and leave the dialog box open. *This button is enabled only when there are differences between the alignment shown in the Alignment Pane and the edited alignment.* The **Cancel** button discards all the changes made since the last use of **Apply**.

### Templates

Template files contain all the Alignment Setup, the Plot Setup, or the Alignment Display Setup properties for the AlignX program. When an alignment project is being created, AlignX locates the default template file (default.xp1 for protein and default.xn1 for DNA) in the main Vector NTI directory. If the file exists, the properties from it apply to the project. If the file is not found, default predefined values from AlignX itself are used.

(To change the default files, a new default.xp1 or default.xn1 file must be created and placed in the top level of the database directory, not the main Vector NTI program directory. Alternatively, additional template files can be retained and reloaded from any directory or path.)

You can save current settings of the alignment project in a template file or apply settings from an existing template to an open project. To save settings as a template, choose **Edit > Setup > Save to File** and select the destination for the settings (or type the name of the new template file). *Default extension for template files is “xp1” for proteins and “xn1” for DNA.* To apply settings from an existing template to the open project, choose **Edit > Setup > Load from File** and select the template file. Settings affecting the plot presentation (Plot Setup) or the alignment presentation (Alignment Display Setup) take effect immediately. The alignment properties (Alignment Setup) take effect only after the alignment is recalculated. To do this, select and realign the same molecules from the Text Pane.

### Identity Table Analysis

The Identity Table analysis shows identity, similarity, and/or distance values for all possible sequence pairs for the sequences that are included in the alignment.
To open the Table, choose **Align > Show Identity Table** from the menu. An window opens like the one in Figure 28.19:

![Similarity Table window](image)

**Figure 28.19 Similarity Table window**

There are two (three for proteins) different pairwise comparison analyses accessible via the buttons on the top of the window.

- Click the top **Identity** button ( ![Identity Button](image) ) to show identity scores between sequence pairs. This is the percentage of identical residues among all ungapped positions between the pairs.
- Click the top **Conserved** button ( ![Conserved Button](image) ) (proteins only) to show similarity values between the sequence pairs. This is the percentage of identical + similar amino acids among all ungapped positions between the pairs.
- Click the top **Distance** button ( ![Distance Button](image) ) to show distance scores (e.g., 100 – identity score) between sequence pairs.

You can also use the **Print Table** button to print the table or the **Camera/Copy Table to Clipboard** button to copy the table to the Clipboard so that it can be pasted into another application, such as Word or Excel.

**Dot Matrix Analysis**

The Dot Matrix analysis is primarily a method for comparing two sequences to find all possible matches of residues. This method can also be used to find direct or inverted repeats in protein and DNA sequences. It can predict regions in RNA that are self-complementary and therefore might form a double-stranded region or secondary structure.

In the Dot Matrix method of sequence comparison, one sequence (A) is listed on the x-axis of the graph and the other sequence (B) is listed on the y-axis. Starting with the first positions in A and B, the program slides the window of n characters along the sequences, performing a comparison of adjacent positions in the windows. If the similarity of residues in each position is above a certain cutoff, a dot is placed in the matrix in the position defined by the starting positions of the window for both sequences. A diagonal line segment indicates that the two sequences match consistently over an extended region.

A larger window size is generally used for DNA sequences than protein sequences since the number of random matches is much greater for DNA.

**Note:** The window size and matching stringency for the Dot Matrix analysis are set in the Dot Matrix Setup dialog box (see *Dot Matrix Setup*, page 412).
Launching the Dot Matrix

To perform a Dot Matrix analysis, open a Dot Matrix window by clicking the **Dot Matrix** button ( ) or by choosing **Align > Show Dot Matrix Plot**. Initially, the Dot Matrix window (Figure 28.20) is empty.

All molecules present in the AlignX window Text Pane are available in both drop-down menus at the top of the empty screen (Figure 28.20). To calculate a Dot Matrix for a pair of molecules, select the molecules in the two drop-down boxes in the Dot Matrix window. A new Dot Matrix is calculated any time a new selection is made in either drop-down menu (the second should not be empty).

![Figure 28.20 Dot Matrix window drop-down menus](image)

The Dot Matrix window contains a menu bar, Main Toolbar, molecule selection toolbars, status bar, and Dot Matrix Pane. Most toolbar buttons have associated tool tips that explain the button function; hover the cursor over the button to display the tool tip for that button. Toolbar buttons are described in this section as needed.

Dot Matrix Window

The Dot Matrix Window is represented in the following example (Figure 28.21):

![Figure 28.21 Dot Matrix Window](image)

The graphical representation of a Dot Matrix is a dashed rectangle in which any region of similar sequence is revealed by a diagonal row of dots. The molecule selected first corresponds to the horizontal axis, its name being shown in the right top corner of the Dot Matrix rectangle. The second molecule corresponds to the vertical axis. Each axis has its own scale that initially displays the sequence positions. When changing the scales (resizing Dot Matrix Window or zooming Dot Matrix Pane) it is possible to show residues along the axes. Residues are displayed only if they can be displayed on both axes. When the scales allow displaying of residues, a grid can be laid upon the Dot Matrix. To apply (or revoke) a grid, click the **Show Grid** ( ) button or choose the **Grid on Zoomed Matrix** command from the shortcut menu or the View menu.
There are two modes in displaying a Dot Matrix, alternated by clicking the Maintain Isotropic Scaling button ( ) or selecting View > Maintain Isotropic Scaling:

Equal scales are used for both axes, meaning the diagonal rows of dots always have a 45° angle relative to the axes.

Various scales for axes, that means that Dot Matrix always occupies maximum space in the Dot Matrix Pane.

The scales for X & Y-axes change dynamically when you resize the Dot Matrix Window. Also, the boundaries of regions currently displayed are shown for both molecules.

Resizing the Dot Matrix

When the Dot Matrix is first calculated, it is shown as a representation of the full length of a molecule. Generally it is impossible to analyze molecules of several thousand residues each when displayed in a 500 x 500 (approximate) screen matrix. To alleviate this problem, a region of a Dot Matrix may be enlarged: click and hold the left mouse button within the Dot Matrix rectangle while dragging the mouse, displaying the tracking dotted rectangle. When the mouse button is released, the selected region is enlarged to the entire Dot Matrix rectangle. The previous state is saved in the history list. Click the Undo Zoom ( ) and Redo Zoom ( ) buttons for modifying the view of the selected region.

Dot Matrix Setup

The window size and stringency (minimum number of matches) for the Dot Matrix analysis is set in the Dot Matrix Setup dialog box. Open the Dot Matrix setup (Figure 28.22) by clicking the Matrix Setup ( ) button on the main toolbar or by choosing the Matrix Setup command from the Matrix menu.

![Figure 28.22 Dot Matrix Setup dialog box](image)

Two parameters affect the Dot Matrix calculation:

<table>
<thead>
<tr>
<th>Dot Matrix Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>stringency (30% by default)</td>
</tr>
<tr>
<td>window (5)</td>
</tr>
</tbody>
</table>

Table 28.7  Dot Matrix setup

Other AlignX Features

The following features can be performed in the AlignX display window. Details specific to AlignX are outlined here. For general information on these features, refer to Chapter 3.

- Print and Print Preview
- Copying to the clipboard
Notes:

Data or graphics from any pane, including the Dot Matrix can be copied to the Clipboard.

When copying the Alignment Pane with the Camera, the Consensus sequence can be excluded from the copy, if desired.

For detailed information on copying the reverse/reverse complement to the Clipboard, see page 398.

- Exporting molecules

Note:

To export a currently open AlignX project into the MSF format, choose Project > Export MSF Format.

- External tools

Alignment Algorithms

For details regarding alignment algorithms, refer to Appendix B.

Alignment PCR

Alignment PCR allows you to design PCR primers for amplifying a region of aligned DNA molecules. To launch an Alignment PCR primers search for a DNA/RNA alignment, in the AlignX Alignment Pane, highlight the region of the alignment you want to amplify, right-click on the Alignment Pane and choose Design PCR Primers from the shortcut menu. The Alignment PCR dialog box (Figure 28.23) displays:

![Alignment PCR dialog box](image)

The Alignment PCR dialog box allows you to set parameters for your alignment PCR primers search. For detailed information on performing Alignment PCR, see Alignment PCR Analysis, page 260.
Introduction

BioAnnotator is a sequence analyzer that performs certain types of DNA and protein sequence analyses, and displays the results as linear graphics. For proteins, amino acid scaling allows the user to visualize the distribution of the particular physiochemical properties along the polypeptide chain, thus providing helpful hints about the functional properties of the protein regions or domains. An amino acid scale is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity, hydrophilicity, and the secondary structure conformational parameters scales, but many other scales exist which are based on different chemical and physical properties of the amino acids. This program provides about 60 pre-defined scales entered from the literature. BioAnnotator also includes the following advanced search and analysis tools, which are run in the Analysis Monitor (see Chapter 34): ProSite, PFAM, BLOCKS and Proteolytic Cleavage.

Launching BioAnnotator

To open BioAnnotator from the Start menu, select Start > Programs > Invitrogen > Vector NTI Suite 10 > BioAnnotator. To launch BioAnnotator from Database Explorer, with a molecule highlighted, select Analyses > BioAnnotator > Analyze Selected Molecule, or select Analyses > BioAnnotator > Open New Analyzer Window. From a Molecule Display Window or other Vector NTI applications, make the same selections, although, of course, the molecule to be analyzed will be that opened in the display window.

You can also drag a molecule from a partial Database Explorer screen or other Vector NTI applications into an empty BioAnnotator workspace.

Features of the BioAnnotator Workspace

Depending upon how BioAnnotator is first opened, the display window may be empty. As a new project is created or an existing BioAnnotator project is opened, the analysis results then appear in the display window.
If your BioAnnotator display window is empty, open a demo project into the workspace before you review its features. Select **File > Open**. Find the Vector NTI Suite folder. It is probably located in Program Files\Vector NTI\Demo Projects. Open the DNA or protein.apf file.

The BioAnnotator interface (Chapter 29.1) consists of a menu bar, three toolbars, and a BioAnnotator Display window, with shortcut menus specific for each pane.

BioAnnotator has three toolbars: the Main toolbar, the Pane Selection toolbar, and the Pane toolbar. The Main Toolbar contains tools for major program functions, such as opening BioAnnotator analysis projects, importing molecule data from various formats, adding and deleting analyses, etc. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

Each Display Window is divided into three section or panes: a Text Pane, a Plot Pane, and a Sequence Pane. Split bars allow the resizing of the panes in the display window. In BioAnnotator, the heavy line at the bottom or top of a graph is also a split bar. Each pane has its own scroll bars, allowing vertical and horizontal movement through the pane.

As in the other Vector NTI Advance applications, to apply a command from the main menu in a given pane (Print, Print Preview, Camera), that pane must be active. The active pane can be toggled with the Switch Panes buttons in the Pane Selection Toolbar or by clicking anywhere in the pane you want to activate.

**Elements of BioAnnotator Display Windows**

**Text Pane**

To activate the Text Pane, click on the **Text Pane** button on the Pane Selection Toolbar. The Text Pane contains a thorough description of the analyzed molecule. The information is organized in a tree structure with the following folders:
Expanding and collapsing folders
Opening the folder shortcut menus
Copying pane contents to the clipboard

Plot Pane

The graphs for a DNA or protein molecule in BioAnnotator display physiochemical properties, many of which are based on parameters like charge that exert effects over distance. Other properties represented in the plot depend on the way adjacent bases/amino acids fold in 3-dimensional space, which is a function of the sequence itself.

The Plot Pane consists of the graphical analyses region: a vertical (Y) axis, showing minimal and maximal values of analysis results, and an individual horizontal (X) axis, displaying either numerical positions in the sequence or residues, scrollbars, and the legend region that displays the name of each analysis. *Even if residues do not show initially, if you enlarge the region enough, individual residues are displayed.* To change the width of the description section, drag the thin vertical bar in the right corner of the common horizontal scale.

At any point along the sequence (X-axis), the value (Y-axis) for the property is derived not just from the specific residue at that point, but from adjacent residues as well. For each property, the algorithm authors have determined the apparent optimum window of adjacent residues to be considered when calculating the value for a point. For instructions in modifying the Window Size parameter, see *Editing Analysis Properties* on page 419.

**Note:** You may note that there are no values for the property at the beginning (or end) of the sequence. Those regions do not have enough bases/amino acids to the left (or right) of the selected residue (as specified by the Window Size) to make the full calculation. To calculate values for those regions, you can reduce the Window Size. Directions for doing so are found in *Editing Analysis Properties* on page 419.

**User Interaction Modes**

BioAnnotator has two *modes* for working with plots.

- *Regular Mode* (default mode): In this mode, regions of data can be selected, the scale of plots can be changed, and the data section can be scrolled. In this mode, the Plot Pane
is synchronized with the Sequence Pane. This means that a selected data region on the Plot Pane is displayed concurrently in the Sequence Pane and vice versa.

- In the Plot Layout Mode: The plots are treated as pictures that can be arranged in a number of ways, including changes in positions and the heights of the plots. To activate the Plot Layout Mode, click the Edit Layout button ( ) on the Plots Pane Toolbar. This enables several edit buttons at the far right of the toolbar.

By default, plots do not overlap, but you can resize them by dragging horizontal separator bars.

**Plot Layout Mode**

To perform any of the following operations, BioAnnotator must be in Plot Layout Mode which is activated by clicking the Edit Layout button ( ).

**Selecting Plots**

To perform certain actions on displayed plots, you must activate the Plot Layout mode by clicking the Edit Layout button ( ), then select one or more plots. To select a plot, click on it with the mouse. If you hold the SHIFT key down, previously selected plots stay selected; otherwise, they become unselected as you click on new selections. Selected plots can be distinguished from unselected ones by the black boxes or “handles” in the corners of the selected plot (Layout Mode only—see below).

**Overlapping Plot Graphics**

- If the Enable Overlapping option is turned on (with the Enable Overlapping button ( ), one plot can be superimposed on another (drag it into position). The overlapping order can also be modified. To bring a plot to the front, use the Bring to Top button ( ) on the toolbar or select View > Bring to Top. To send a selected plot to the back, press the Send to Bottom button ( ) on the toolbar or select View > Send to Bottom.

- When the Enable Overlapping option is turned off, superimposed plots return to their original order.

**Changing Plot Position**

- If the Enable Overlapping option is on, the plot can be dragged and dropped anywhere in the Plot Pane. Click on a plot + drag the mouse. The cursor shape is changes from  to  . It will be placed on top of other plots that might occupy the same position.

- If the Enable Overlapping option is off, the plot can be dropped in specific positions only.

**Changing Plot Height**

Plot height can be changed only for a single plot at a time.

- To modify the plot height, move the cursor to one of the handles where it changes to a two-headed arrow ( ). Drag the handles vertically to resize the plot.

- If the Enable Overlapping option is on, handles appear at the top and bottom of a graph.
Regular Mode

Selecting Data
A region of data may be selected in the Regular Mode only. To select a region, press and hold the mouse button and drag the mouse horizontally over the plot. To modify the current selection either hold the SHIFT key down and click/drag in the plot area or resize the edges of the existing selection by dragging them.

By clicking outside the plot area or individual horizontal axis, BioAnnotator removes the selection from plot.

Selecting a region on one of the plots also changes selected regions on all plots.

Changing Scale
As in other Vector NTI Advance applications, plot analysis graphs can be resized using the Zoom buttons. The scale for all plots can be changed by using the **Zoom In** ( ) and **Zoom Out** ( ) buttons or by selecting **View > Zoom In** or **Zoom Out**. Increase the scale of a selected region in the Plots Pane by clicking the **Zoom to Selection** button ( ). In this case, now you can read the individual residues in the Plot Pane. To fit the entire plot into the data section, use the **Fit to Window** button ( ) or select **View > Fit to Window**.

Editing Analysis Properties
To modify individual properties and common display properties, double-click on a plot, opening the Plot Properties dialog box. Alternatively, click on a plot once to select it, then press the **Properties** button ( ) or choose **View > Properties** command from the menu bar.

In the Plot Properties dialog box, both the analysis parameters and the display setup for the plot can be changed.

- The Analysis tab displays a short description of the analysis as well as the Window Size parameter, which is analysis dependent. The Window Size is the number of adjacent data points used in averaging values for each displayed data point. (See the description of Window Size under **Plot Pane** on page 417.) You may alter the Window Size to reflect experimental findings or thinking that has changed since the author first reported the algorithm, or you may want to calculate residues within a blank zone at the beginning or end of the sequence. Try varying the window size to see how it affects the plot display.

- The Datagram tab changes the display attributes of the plot, such as plot type and color. It also turns the display of the individual scale for the plot on or off.

Changing Common Display Properties
The common display properties of all plots can be customized using buttons on the Plot Pane toolbar of the View menu. The **Show Vertical Axis** command/buttons ( ) turns on/off the display of vertical axes on all plots. The **Show Legends** ( ) and **Place Legends Below Plots** ( ) buttons turn on/off and change position of the legends for all of the plots.

Display Setup for Analyses is discussed later in this chapter.
Sequence Pane

Like other Vector NTI Advance applications, in BioAnnotator, nucleotide or amino acid sequences can be viewed, manipulated and formatted in the Sequence Pane. Regions of the sequence can be copied and pasted to the Clipboard.

Sequence Pane Elements

In addition to the sequence, the Sequence Pane displays a molecule’s sequence features. Features are shown with a horizontal line above the sequence (protein features and DNA features on the direct strand) or below the sequence (DNA features on the complementary strand). Labels also show the name of the feature.

The sequence can be scrolled up or down with the Sequence Pane’s scroll bar. If Sequence Pane is the active pane, standard arrow keys may be used to scroll the sequence.

Navigating the Sequence Pane Using the Keyboard

Standard keyboard keys may be used to navigate the sequence:

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT</td>
<td>Move caret to previous position</td>
</tr>
<tr>
<td>RIGHT</td>
<td>Move caret to next position</td>
</tr>
<tr>
<td>UP</td>
<td>Move caret position one line up</td>
</tr>
<tr>
<td>DOWN</td>
<td>Move caret position one line down</td>
</tr>
<tr>
<td>HOME</td>
<td>Move caret position to the beginning of the current line</td>
</tr>
<tr>
<td>END</td>
<td>Move caret position to the end of the current line</td>
</tr>
<tr>
<td>CTRL-PgUP</td>
<td>Move caret position to the beginning of sequence</td>
</tr>
<tr>
<td>CTRL-PgDN</td>
<td>Move caret position to the end of sequence</td>
</tr>
</tbody>
</table>

Table 29.2 Navigating a sequence using the keyboard

Selecting a Sequence Region

To select sequence regions, CLICK + DRAG the mouse across the sequence. Press SHIFT + RIGHT or LEFT ARROW to start or extend a selection.

A selected region in the Sequence Pane is concurrently displayed on the plots in the Plot Pane.

Changing Sequence Pane Properties

In BioAnnotator, double-strand mode is default mode for DNA/RNA sequences. To switch to a single strand display, press the ( ) button on the pane toolbar or select View > Show Both Strands.
The length of a single line of a sequence and number of blocks per line can be adjusted as well. Click the **BioAnnotator Setup** button on the toolbar to open the Setup dialog box (Figure 29.2):

![Setup dialog box](image)

**Figure 29.2 Setup dialog box**

Click the **Setup** button for a **Sequence** to control sequence display (Chapter 29.3):

![Sequence Display Setup dialog box](image)

**Figure 29.3 Sequence Display Setup dialog box**

To choose what features should be shown in the Sequence Pane, click on the **Setup** button for **Feature Map** in the Setup dialog box, opening the FMap Setup dialog box (Chapter 29.4):

![FMap Setup dialog box](image)

**Figure 29.4 FMap Setup dialog box**
Click the <Add> button to add features for display. Feature Map Setup is discussed in detail in Editing the Feature Map on page 240.

Changing Sequence Character Display Attributes
Display attributes for selected sequence residues can be changed in the same way that text is formatted in a word processor using the appropriate toolbar buttons. See Chapter 3 for details.

Copying a Raw Sequence
To copy real sequence data to the Clipboard, select a sequence region and choose Edit > Copy. The sequence will be copied in the plain text format without formatting or line numbering. This “raw” format is suitable for pasting into other molecular biology software including programs in Vector NTI Advance. The keystroke for the copy operation is CTRL+C.

BioAnnotator Projects

Creating a New Project
When BioAnnotator is launched, a new, empty display window is opened. To create a new project, you must import molecule information from various molecule file formats. To import a molecule, select File > Open on the menu bar.

BioAnnotator reads the following file formats: GenBank and EMBL for DNA, GenPept and SWISS-PROT for proteins. BioAnnotator also reads ASCII Text, FASTA, and Vector NTI archives for both DNA and proteins.

In the File Open dialog box that opens, select the desired file and click Open. The program tries to recognize which format is being used and whether the sequence is nucleic acid (DNA/RNA) or amino acid (proteins).

If the program fails to define a file’s format, it will be suggested in the following dialog box that the file be imported as a raw sequence (Chapter 29.5):

![Figure 29.5 Importing a raw sequence](image)

Figure 29.5 Importing a raw sequence

To create a new BioAnnotator project from Vector NTI, in Database Explorer, select a molecule, then choose Analyses > BioAnnotator > Analyze Selected Molecule or Analyses > BioAnnotator > Open New Analyzer Window.

Opening an Existing Project
BioAnnotator can save a molecule and its an analysis list (with individual options) and displayed character attributes, as a single file. At any time, BioAnnotator can restore the contents of this file and continue to work with it. An aggregate set of molecule, options, and attributes is called a BioAnnotator project.
To open an existing project, choose **File > Open** or press the ( ) button on the toolbar. In the Open dialog box, select the desired file and click **Open**. **By default, BioAnnotator project files have an ".apf" file extension.** If the file is an analyzer project, information about the molecule, its analyses, and sequence display character properties (if any) appears in the appropriate panes. Another way to open an existing project is to drop the project file on the BioAnnotator window.

### Analysis List Setup

The important part of configuring BioAnnotator is the selection of analyses to perform for a molecule. To add or remove any analysis from the plot, open the Analysis List Setup dialog box where you can choose from the list of available analyses.

To open the dialog box, click the **Analysis List** button ( ) on the Main Toolbar, choose **Analyses > Analyses List** on the menu bar or click the **Setup** button for **Analysis List** in the **BioAnnotator** Setup dialog box described previously in this chapter. This opens the Analysis List Setup dialog box (Chapter 29.6).

![Figure 29.6 Analysis List Setup dialog box](image)

The dialog box lists currently displayed analyses in the Selected Analyses box and a list of 50 predefined analyses in the Available Analyses box. **Available analyses differ for DNA/RNA and protein molecules.** Citations for each analysis are noted in the dialog box.

To add one or more analyses to the displayed analysis list, in the upper list, select the appropriate analyses and click the ( ) button or double-click on each analysis to be added. To see the same analysis with different parameters, you can add more than one “instance” of the same analysis.

The relative position of an analysis in the Selected Analyses list determines the order in which the plots appear in the Plot Pane. The list may be rearranged by moving a selected analysis up or down by clicking ( ) and ( ) buttons.

Once analyses appear in the bottom box, properties of the selected analyses can be changed, analyses may be moved up and down in analysis list, and analyses may be removed from the list. To change the properties of an analysis, select it in the list and click the ( ) button or double-click the analysis in the list.
To remove an analysis from the displayed analysis list, select it in the analysis list and click the ( ) button.

**Analysis Properties**

To change analysis parameters and choose from the various plot representation options, open the Analysis Properties dialog box by clicking the ( ) button in the Analysis List Setup dialog box, or the same button on the Plots Pane Toolbar or by double-clicking on a plot in the Plot Pane.

The Analysis Properties dialog box has two tabs, the Analysis tab and the Datagram tab.

**Analysis Tab**

On the Analysis tab, you can specify analysis-dependent analysis parameters. In general, every analysis has its own version of this tab for editing analysis-specific parameters. However, most protein analyses have very similar Analysis tabs. Examples of the Analysis tab for two different analyses are described below.

1. The Analysis tab for the ProtScale analyses has the following format (Chapter 29.7):

   ![ProtScale analyses: Analysis tab of the Setup dialog box](Figure 29.7 ProtScale analyses: Analysis tab of the Setup dialog box)

   o **Minimal Value** – Minimum value considered for weighted average.
   o **Maximal Value** – Maximum value considered for weighted average.
   o **Window Size** – Number of nucleotides or amino acids used to calculate average/weighted average value, which is plotted on the graph. If you hover your cursor over a point on a graph, a tool tip appears showing the number of residues in the window (set window size) and the value calculated for that point using the set window size.
The Analysis tab for the Sequence Complexity analysis has the following format (Chapter 29.8):

![Figure 29.8 Sequence Complexity analysis Analysis tab](image)

**Datagram tab**

This Datagram tab is the same for all analyses (Chapter 29.9).

![Figure 29.9 Analysis Properties dialog box](image)

In the Personal Graph Properties section, you can specify the following Datagram properties:

- Show Individual Scale – Check/uncheck the box to display/hide the individual horizontal axis.
- Visual Type – Choose the plot type (Bar Chart, Line Chart, Boolean Chart).
- Threshold – The minimal relative value considered as “true” for Boolean Chart.
- Color – The color of the graph.

The datagram configuration is shown in the preview box on the tab. Choose **OK** to apply any changes you have made.
Analysis Descriptions

DNA/RNA Analyses

Melting temperature and free energy are calculated using the “nearest neighbors” method. Constants and algorithms used for calculation of thermodynamic parameters may be found in the sources listed in Appendix D.

The full list of DNA/RNA analyses are as follows:

- Free Energy (dG) (kcal/mol)
- Enthalpy (dS) (kcal/mol)
- Entropy (dS) (cal/K/mol)
- Melting Temperature (Thermodynamic) ©
- Melting Temperature (GC Content) ©
- GC Content (%)
- Nucleic Acid Distribution (%)
- Sequence Complexity

Protein Analyses

Nearly all protein analyses are based on ProtScale analyses. ProtScale allows the computation and representation of the profile produced by any amino acid scale on a selected protein.

An amino acid scale is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are the hydrophobicity or hydrophilicity scales and the secondary structure conformational parameters scales. Many other scales exist which are based on different chemical and physical properties of the amino acids.

The full list of ProtScale analyses is as follows: Full citations appear in Appendix B.

- Amino acid composition (%) in the SWISS-PROT Protein Sequence data bank
- Antigenicity value X 10
- Atomic weight ratio of hetero elements in end group to C in side chain
- Average area buried on transfer from standard state to folded protein
- Average flexibility index
- Average surrounding hydrophobicity
- Bulkiness
- Conformational parameter for alpha helix (computed from 29 proteins)
- Conformational parameter for alpha helix
- Conformational parameter for beta-sheet (computed from 29 proteins)
- Conformational parameter for beta-sheet
- Conformational parameter for beta-turn (computed from 29 proteins)
- Conformational parameter for beta-turn
- Conformational parameter for coil
- Conformational preference for antiparallel beta strand
- Conformational preference for parallel beta strand
- Conformational preference for total beta strand (antiparallel+parallel)
- Free energy of transfer from inside to outside of a globular protein
- Hydration potential (kcal/mole) at 25 C
- Hydropathicity
- Hydrophilicity scale derived from HPLC peptide retention times
- Hydrophobic constants derived from HPLC peptide retention times
- Hydrophobicity (delta G1/2 cal)
- Hydrophobicity (free energy of transfer to surface in kcal/mole)
- Hydrophobicity indices at pH 3.4 determined by HPLC
- Hydrophobicity scale (contact energy derived from 3D data)
- Hydrophobicity scale (pi-r)
- Hydrophobicity scale based on free energy of transfer (kcal/mole)
- Mean fractional area loss (f) (average area buried/standard state area)
- Mobilities of amino acids on chromatography paper (RF)
- Molar fraction (%) of 2001 buried residues
- Molar fraction (%) of 3220 accessible residues
- Molecular weight of each amino acid
- Normalized consensus hydrophobicity scale
- Normalized frequency for alpha helix
- Normalized frequency for beta-sheet
- Number of codon(s) coding for each amino acid in universal genetic code
- Optimized matching hydrophobicity (OMH)
- Overall amino acid composition (%)
- Polarity (p)
- Polarity
- Proportion of residues 95% buried (in 12 proteins)
- Recognition factors
- Refractivity
- Relative mutability of amino acids (Ala=100)
- Retention coefficient in HFBA
- Retention coefficient in HPLC, pH 2.1
- Retention coefficient in HPLC, pH 7.4
- Retention coefficient in TFA

**ProtScale Algorithm**

All of the ProtScale analyses use one algorithm:

for average method and

for Weight Method.

where,

\[ D_i \] \(-\) plot value in \( i^{th} \) position

\[ W \] \(-\) window size

\[ S_i \] \(-\) amino acid in \( i^{th} \) position
\[ D_i = \frac{\frac{W}{2}}{\sum f(S_{i+j})} \]

\[ D_i = \frac{\frac{W}{2}}{\sum f(S_{i+j}) \times W h_j} \]

\( f() \) - coefficient for amino acid from table of coefficients (specified for each of analysis)

\( W h_j \) - linear weight coefficient

**Advanced Analyses: BLOCKS, ProSite, PFAM and Proteolytic Cleavage**

To launch an advanced analysis for the protein molecule displayed in the BioAnnotator workspace, choose **Analyses > BioAnnotator > <analysis type>** from the menu. Analysis Monitor opens for the molecule with the appropriate parameters interface for the chosen analysis. For more information on performing advanced analyses in Analysis Monitor, see Chapter 34.

**Note:** The advanced analyses available in BioAnnotator also become available from the Vector NTI Molecule Viewing window, assuming both BioAnnotator and Vector NTI are licensed.

**Adding Custom Protein Analyses**

Specific values are used in the calculation of the various protein analyses. These values differ depending on the particular analysis and are stored in small text files having a .pa2 extension. All .pa2 analysis files are stored in the Analyses subdirectory in the Vector NTI Suite directory. The name of the analysis corresponds with the name of the .pa2 file.

For example, the Bulkiness analysis is characterized in the Bulkiness.pa2 file as follows:

```plaintext
[general]
author=Zimmerman J.M., Eliezer N., Simha R.
```
The [general] section describes the attributes of the analysis and the [data] section contains the list of values used for calculating the analysis. Values are listed according to the alphabetical order of the acid coefficients.

You can create your own custom protein analysis by modifying the values in one of the existing .pa2 files using a plain text editor, such as Notepad. Once modified, the new .pa2 file should be saved in the Analyses subdirectory under a new name that corresponds to the new analysis type. If BioAnnotator is open at the time the new .pa2 file is created, it must be closed and reopened before the new analysis is recognized.

Other BioAnnotator Features

The following features can be performed in the BioAnnotator Display window. Details specific to BioAnnotator are outlined here. For more information, refer to Chapter 3.

- Printing from BioAnnotator
- Copying data from BioAnnotator
- Exporting molecule data
Introduction

ContigExpress is a program for assembling and editing sequencing fragments, either in the form of text sequences or chromatograms from automated sequencers, into longer contiguous sequences or “contigs.”

ContigExpress uses CAP3 (Huang and Madan, 1999; Huang, 1996) to drive the assembly process. This widely-used sequence assembly program can use quality value scores (QVs) in ends trimming, contig construction and consensus calculation. It also produces excellent results when QVs are unavailable. It is capable of using forward-reverse constraints to evaluate contigs, a feature that helps in accurate placement of repetitive sequence fragments. It can also identify and discard chimeric sequencing reads that frequently result from lane tracking errors. One of the major strengths of CAP3 is its consensus generation algorithm based on weighted sum of QVs.

Work in ContigExpress takes place in a ContigExpress Project, a file where you store fragments, their assemblies, and assembly options related to your current task. In ContigExpress, fragments can be edited directly, with the chromatograms in full view. Changes are tracked and a history is maintained. The contigs generated are then saved and managed as GenBank, EMBL or FASTA files. They can be exported into other applications of Vector NTI Advance for further analysis or visualization.

Launching ContigExpress

To launch ContigExpress from the Start menu, select Start > Programs > Invitrogen > Vector NTI Suite 10 > ContigExpress. To launch ContigExpress from Vector NTI Database Explorer or other Vector NTI Advance applications, select Assemble Menu > ContigExpress – Assemble Selected Molecules or Assemble Menu > ContigExpress – Open New Assembly Project.
Features of the ContigExpress Workspace

ContigExpress consists of three component windows:

- **ConfigExpress Project Explorer**: This window is the first window you see when you launch ContigExpress. In this window you open, save, and close the projects and set the assembly options. Project Explorer is also where you can import, organize, and assemble your sequencing fragments.

- **Fragment Viewer**: This is the workspace where you work with individual fragments and chromatograms.

- **Contig Window**: Contig Window is the place where you review and edit contigs.

*Note: Fragment and Contig Windows are covered in Chapter 31.*

**CE Project Explorer**

Project Explorer, similar to the Vector NTI Database Explorer, is central to the management of sequencing projects and assembly of fragments into contiguous sequences. When first opened, the project window is empty. The screenshots in this and the following two chapters use sequencing fragments supplied with your installation. Select **Project > Add Fragments > From ABI file** from the Project window menu. In the Open dialog box, navigate to folder HS354_C02 (located in C:\Program Files\Invitrogen\Vector NTI Suite 10\Demo Projects if a default Vector NTI Advance 10 installation has been performed). Open the folder and the ABI subfolder inside, select all the fragments and click the **Open** button.

**Elements of Project Explorer**

The CE Project Explorer window (Figure 30.1) consists of a menu bar, a toolbar, a Tree Pane on the left and a List Pane on the right side of the window.

![Figure 30.1 CE Project Explorer interface](image)

As you are introduced to various operations in the chapters covering ContigExpress, the toolbar buttons are displayed as needed. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.
The ContigExpress Project Explorer window is divided into two panes: a Tree Pane on the left and a List Pane on the right. As a new project is created or an existing ContigExpress project is opened, the fragment and Assembly folders are listed in the Tree Pane. In Contents View, folders can be opened (expanded) or closed (collapsed) as in the other applications of Vector NTI.

Status bars in Project Explorer show information about the items in the Tree and List panes.

The status bar above the display panes shows the current Tree Pane view mode and the number of items contained in the selected project in the Tree Pane (Figure 30.2).

The status bar at the bottom of the Project Explorer shows the amount and type of currently selected item(s). It can be configured to show one of several properties of a selected single item in the List pane (View menu > Options > Status Bar tab, or left mouse-click on column header).

### Tree Pane

The Tree Pane displays the general structure of your ContigExpress Project in either of two viewing modes: History or Content view. The contents of each pane are displayed and manipulated similar to other Windows applications, such as Vector NTI Database Explorer or Windows Explorer.

The following types of items can be found in the Tree pane:

- ![List all fragments in the project](image)
- ![Assembly of fragments](image)
- ![Contig](image)
- ![Fragment included in an assembly](image)
- ![Fragment not included in an assembly](image)
**History View**

![History View Diagram](image)

The root of the tree is always the list of all fragments existing in the project. The tree nodes are assemblies that have descendant assemblies, and the leaves are the assemblies that have no descendants.

**Content View**

Click the Content View button ( ) to display a “flat” list of all assemblies in the project without displaying their mutual relationships. In Content View, you can display in the Tree Pane the contigs in the assemblies and non-included fragments by clicking on the + button at the left of assembly name.

As in the History View, the root of the tree always list all of the fragments existing in the project. The tree nodes are the assemblies. Contigs and fragments that were attempted to be included but were unsuccessful are listed under the assemblies. Fragments successfully included in the assemblies are listed inside the Contig Folders (Figure 30.4).

![Content View Diagram](image)

**Table 30.1  Viewing modes**

<table>
<thead>
<tr>
<th>Viewing Mode</th>
<th>Assemblies</th>
<th>Contigs</th>
<th>Fragments in Assemblies</th>
<th>Fragments not Included in Assemblies</th>
</tr>
</thead>
<tbody>
<tr>
<td>History View</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Contents View</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Figure 30.3 Tree Pane History view

Figure 30.4 Contigs and fragments that were not successfully included are listed under the assemblies
List Pane

The List Pane shows the contents of an item currently selected in the Tree pane. When the Fragments (MAIN) is selected in the Tree Pane, the List Pane displays all fragments in the project (Figure 30.5).

![Figure 30.5 List Pane displays contents of item selected in Tree Pane](image)

When an assembly is selected in the Tree Pane, the List Pane contains:

- the contigs contained in the assembly
- the fragments that comprise contigs
- fragments that were considered in the assembly process, but were not included in the final assembly.

The following types of items can be found in a List Pane (see Figure 30.2). Note that the symbols to the left of each list item identifies its place in the project:

- ![Contig](image)
- ![Lite Contig](image) A contig generated under Lit& Contig mode, in which imported chromatogram data is no longer viewable (page 450).
- ![Fragment not considered in an assembly](image)
- ![Fragment included in a contig](image)
- ![Fragment considered in an assembly but not used in a contig (grayed out symbol)](image)

A fragment not included in a contig contains no other items; therefore, nothing is shown in the List pane when it is selected in the Tree pane.
A contig contains its fragments (Figure 30.6):

![Figure 30.6 A contig contains its fragments](image)

A fragment not included in a assembly contains no other items; therefore, nothing is shown in the List pane when it is selected in the Tree pane.

On the toolbar, the List Format buttons control the list display. In the first three modes, only item icons and names are displayed.

### List Pane Formats

<table>
<thead>
<tr>
<th>Button</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Large Icons" /></td>
<td>View the List pane in Large Icons mode</td>
</tr>
<tr>
<td><img src="image" alt="Small Icons" /></td>
<td>View the List pane in Small Icons mode</td>
</tr>
<tr>
<td><img src="image" alt="List" /></td>
<td>View the List pane in List mode</td>
</tr>
<tr>
<td><img src="image" alt="Details" /></td>
<td>View the List pane in Details mode. In this mode, you can select the properties shown in column for each item type.</td>
</tr>
</tbody>
</table>

**Table 30.2  List Pane format options**

### List Pane Filters

The next group of buttons filters data displayed in the List pane.

<table>
<thead>
<tr>
<th>Button</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Show Contigs" /></td>
<td>Enable the <strong>Show Contigs</strong> filter in List pane</td>
</tr>
<tr>
<td><img src="image" alt="Expand Contigs" /></td>
<td>Enable the <strong>Expand Contigs</strong> filter in List pane</td>
</tr>
<tr>
<td><img src="image" alt="Show Unassembled Fragments" /></td>
<td>Enable the <strong>Show Unassembled Fragments</strong> filter in List pane</td>
</tr>
<tr>
<td><img src="image" alt="Show Other Fragments" /></td>
<td>Enable the <strong>Show Other Fragments</strong> filter in List pane</td>
</tr>
</tbody>
</table>

**Table 30.3  List Pane filter options**
Column Dialog

A Column Dialog (Figure 30.7) can be called up by a right-click on the column header bar:

![Column Dialog](image)

Figure 30.7 Column dialog for configuring appearance of List Pane

The List Pane column fields can be selected or de-selected in the Details View Tab. Order and alignment can be adjusted accordingly. Standard display fields are: Name, Length, Original Length, 5’ Trimmed Bases, and 3’ Trimmed Bases.

The value of “Minimum number of overlapping fragments to calculate unambiguity” dictates that unambiguity calculation algorithm consider only those segments of contig that have more then the specified number of overlapping fragments.

Sorting List Pane Columns

Objects are described in the List Pane by fields in columns. Sort column objects by clicking on the property column headings (in Details mode). Reverse the order by clicking on the same column header for the second time or by selecting View > Arrange Icons.

Note that when contigs are constructed in the Full Contig mode, all component fragments can be opened in the Fragment Window from within List Pane. However, when in the Lite Contig mode, the component fragments cannot be opened because the linkage between the fragments and contigs are not retained. The choice between Full Contig vs Lite Contig mode can be selected in the Lite Settings Tab in the Assembly Setup dialog box.

General Project Explorer Operations

Opening or Creating a Project

Opening an Existing ContigExpress Project

To open a ContigExpress (.cep) project from a disk or other location, select Project > Open Project. Click the Open Project button or use CTRL-O. You can also drag and drop a project file from a folder in Windows Explorer into the ContigExpress Project workspace.
To open a ContigExpress (.cep) project from a disk or other location, select **Project > Open**

**Project**, click the **Open Project** button ( ) or use CTRL-O. You can also drag and drop a project file from Database Explorer into ContigExpress.

In the Open Project dialog box, browse the folders available on your computer and select an existing project file (extension .cep). Select the project file and click **Open**:

![Open project dialog, browsed to VectorNTI Demo Projects folder](image)

If you get an error message, the selected project file probably has an invalid structure. If a project is currently open in the CE Project Explorer and it has been changed since its last save, you will be asked at this point to save the project. If you choose not to open a new project at this stage, be click **Cancel**.

For your convenience, the four most recently opened projects are listed at the bottom of the Project menu option. Select **Project > <#><previously opened project name>** to open one of those listed.

After successful loading, the panes of the Project Explorer are updated to show the new project.

**(Importing and Exporting in ACE Format)**

ContigExpress can import an assembly project in ACE format. In a blank ContigExpress Project workspace, select **Project > Add Fragments > From Phrap ACE file** and choose the ACE file in the standard Windows file open dialog.

To export in ACE format an assembly from a ContigExpress project, select the assembly of interest in the Tree Pane. Select “Export Assembly to ACE” item from the context sensitive shortcut menu, invoked with a right mouse-button click. The name of the ACE file for the assembly can be specified in the file open dialog.

**Importing Fragments to a Project**

To import a fragments to a project, from the CE Project window menu, select **Project > Add Fragments > From <type> file**... or from the shortcut menu launched by right-clicking on **Fragments (MAIN)** in the Tree Pane of the CE Project window, select **Add fragments > From <type> file**. Select the fragment file names in the Import Sequence From dialog box and click **Open**.

Alternatively, select one or several files/fragments in the Windows Explorer or Vector NTI Explorer and drag them to ContigExpress Project Explorer and drop them to any of the Project Explorer panes. Upon dropping, they are added to the project. Similarly, you can use copy and paste.
When importing a large number, say 1000, of fragments from individual files, you should select **Project > Add Fragments > Folder of Fragments.** This helps you to get around a limit set by Microsoft Windows on the number of files one can open simultaneously.

**Accepted Fragment File Types**

Files must be in one of the following supported file formats:

- GenBank (*.gb)
- FASTA (*.txt)
- ABI (*.abi)
- AB1 (*.ab1)
- EMBL (*.txt)
- Staden SCF (*.scf)
- ALF (*.alf)
- plain text sequence (*.txt)
- MegaBASE (.esd)
- Phred (.seq)
- Phrap ACE (.ace)

(Note that if sequence files were passed through an application that renamed them using proprietary file extensions, they may not be visible in their directory when specifying a supported extension suffix (e.g., .ab1). In most cases, simply change the Files of Type setting in the Import dialog to All Files (*.*)).

The imported fragment(s) are added to the project and the list of all fragments is updated.

Sequence chromatograms can be imported through AB1, SCF and ESD formats. Displaying chromatograms during review and editing is obviously preferable. However, when dealing with a large project, say, with more than a few thousands of fragments, computer resources may become a limiting factor. To help meet this challenge, ContigExpress provides two alternatives: one is to discard the chromatograms upon importing; the other is to assemble using Lite Contig mode instead of the Full Contig mode.

Discarding chromatograms can be selected with the “Discard chromatogram data on import” check box in the Lite Settings tab of the Assembly Setup dialog box. This will save about 50 K of memory on average for each read of 500 bases length.

In Lite Contig mode, chromatograms associated with the fragments are discarded and edits performed on the fragments contained in contigs are not applied to the individual fragment files as in the Full Contig mode. Usage of Lite Contig mode reduces the demand for computer memory tremendously and is thus recommended for assembling large number of fragments. ContigExpress provides a warning message when you try to assemble more than 500 fragments in Full Contig mode because of the projected requirement for computer resources during contig editing, if all fragments are assembled into one contig. Note that this number is based on our rough estimation and is definitely not a magic one and can vary in individual installations. It does NOT mean that ContigExpress cannot assemble more than 500 fragments in Full Contig mode. Users are recommended to experiment initially with this and find a number appropriate under your context. When an appropriate number is defined, you can set the controls on Assembly Setup > Lite Settings tab so that ContigExpress operates in Full Contig mode, but automatically switches to Lite contigs when the number is exceeded.

For Lite Contig mode, please refer to **Lite Settings Tab** on page 450.
Finding Fragments

Launch the Find Fragment dialog box by choosing **Edit > Find Fragment** from the Project Explorer menu or by pressing **CTRL+F** on the keyboard. The Find dialog box displays (Figure 30.9):

![Find Fragment dialog](image)

*Figure 30.9 Find Fragment dialog*

Type the object name in the Object Name box and set the search parameters:

- **Match whole word only** – If checked, finds the object only if the entire object name is entered in the Object Name box.
- **Match case** – If checked, finds the object only if the entire object name is entered in the Object Name box in case-sensitive fashion.
- **Search area** – Specify the area of the project you want to search: the assembly currently highlighted only, from the current assembly forward, or the entire project.
- **Advanced search options** – Allows you to search fragments only and/or fragments/contigs of a specified length only. Once you have entered the object name and set the
search conditions appropriately, click the **Find** button to begin the search. Search results display in the lower portion of the Find dialog box (Figure 30.10):

![Find Fragment dialog box](image)

Double-click on a fragment name in the results section of the Find dialog box to highlight the fragment in its assembly in the Project Explorer window.

**Importing Phred Quality Values**

You can load Phred quality values for fragments into your project. In ContigExpress, these values are used in sequence clipping calculation before assembly (see below), consensus determination (see Chapter 31, Contig Viewer Viewing options) or in fragment trimming (see Chapter 32). For more information on Phred quality values, see [http://www.phrap.com/phred](http://www.phrap.com/phred).

To load Phred quality values for existing fragments, right-click on a fragment listed in the Fragments Pane of the Project window and choose **Load Phred Quality Values** from the shortcut menu. In the Open dialog box, navigate to the Phred Quality Score File (.qual), highlight it and click the **Open** button.

AB1 and SCF files may contain QVs. Upon import, these QVs are parsed into ContigExpress. When there are QV files named as xxx.ab1.qual or xxx.scf.qual, the QVs from the .qual file replace the QVs from the corresponding .ab1 or .scf files in the ContigExpress project. Fragment names must match between the .qual and sequence file and the pair must contain an equal number of elements.

**Importing a Directory of Fragment Files**

You can import a directory containing multiple sequencing files (in supported file formats) using the **Import Folder of Fragments** command. To execute this command, choose **Project > Add**
**Fragments > Import Folder of Fragments** from the menu in the ContigExpress Project window.

**Importing FASTA Files**

You can import one or more, or a whole folder of FASTA files into a ContigExpress project. If there are .qual files in the same folder, ContigExpress automatically imports the corresponding Quality Values (QVs) based on the matching file names. For example, if the sequence file is named abc.fsa, then abc.fsa.qual will be imported along with the corresponding sequence file (abc.fsa) and values assigned to the respective bases.

For multi-sequence FASTA files, the corresponding file with a .qual suffix will be recognized just as the single-sequence files. Values are assigned to the respective bases in each of the fragments. One extra requirement is that the first word between the “>” at the line start and the first white space on each header line has to match for each sequence, in the sequence file and .qual files.

**Selection Techniques**

Selection techniques described here are unique to ContigExpress. General selection techniques are described in Chapter 3.

**Tree Pane**

In the Tree Pane, click on an item to select it. (Only one item can be selected at a time.) Change the selection by clicking the item with the left mouse button or by moving the selection bar with the cursor buttons on the keyboard.

**List Pane**

The List Pane allows multiple item selections:

- To select a range of items: SHIFT + CLICK on the first and last item in the list
- To select non-contiguous items: CTRL + CLICK on each item
- To select a group of contiguous items in the List Pane, click anywhere in the blank area of the pane. Drag a box around the files you want to select.
- To select all the items in the List Pane, right-click on any item in the list and choose Select All from the Edit menu, the shortcut menu, or use the keystroke CTRL + A.

When you select an item in the Tree Pane, the contents of the List Pane are updated to show the contents of the newly selected item. The upper status bar is updated as well. The lower status bar reflects selection changes in both panes of the Project Explorer.

To invert a current selection in the List pane (that is, make all currently unselected items selected and vice versa), select **Edit Menu > Invert Selection**.

**Assembling Selected Items**

Select at least two items in the List Pane, and choose Assemble > Assemble Selected Fragments, or click the Assemble Selected button ( ).

If one or several contigs are selected for assembly, the sequences of all fragments that comprise the contig(s) are used as fragment sequences for the assembly process.
Since assembling is a lengthy process, a progress dialog box (Figure 30.11) appears showing you the assembly steps. You can abort the process at any time by clicking the **Cancel** button:

![Assembly progress](image)

*Figure 30.11 Assembly in progress*

If a contig cannot be created from the selected fragments with the selected setup options, you will be informed.

If the assembly succeeds, a new “child” assembly is created inside the current assembly. The selection in the Tree Pane moves to the newly created assembly and the List Pane shows its contents. At the end of the assembly, an Assembly Completion message displays the duration of the assembly, the number of contigs constructed and the total number of fragments used.

**Assembly of a Large Project:**

Although ContigExpress does not have an exact upper bound for project size, your project size may be limited by the available computer resources in terms of both hardware and software (such as operating system). When you attempt to assemble more than 500 fragments in Full Contig mode, ContigExpress generates a warning message to bring this matter to your attention (Figure 30.12).

![Warning](image)

*Figure 30.12 Assembling a large contig in Full Contig mode places heavy demands on graphics units resources and makes editing in the Contig Viewer difficult.*

If you still want to proceed with the current assembly in Full Contig mode, click the OK button. (If you check “Do not show me this message again” ContigExpress will not generate this warning message again.) If you do encounter an “out of memory” situation, you should consider assembly in Lite Contig Mode (**Lite Settings Tab** on page 450). Many a times the “out of memory” problem occurs due to the presence of too many assemblies in the ContigExpress Project. In this case, you are advised to dissolve some of these assemblies, save the project and restart ContigExpress. Limiting the number of assemblies in a project is always a good idea with large projects. If you decide to use Lite Contig Mode but don’t want to lose the linkage between your
contigs and fragments, you can discard chromatograms (page 451) and trade the ability to invoke chromatograms for a larger capacity.

**Setting Assembly Setup Parameters**

When you create or open a Contig Express project, a default set of assembly parameters is used for assembling contigs in the project. To change these values, click the **Assembly Setup** button ( ) or select **Assemble > Assembly Setup**, opening the Assembly Setup dialog box (Figure 30.13):

![Assembly Setup dialog box](image)

**Figure 30.13 Assembly Setup -- Assembly tab**

**Loading and Saving Assembly Setup Settings**

Click the Load and Save buttons to read and save any set of parameters from Assembly Setup. Dialog boxes open allowing navigation to an existing parameter file to be loaded, or to a folder where a new parameter file will be saved. After loading an existing parameter file, all fields of the dialog box are updated.

**Assembly Tab**

There are three main check boxes to control some important aspects of the sequence assembly process: Use Quality Values, Detect Chimeric Reads and Use Forward-Reverse Constraints. These options are explained in Table 30.4.

<table>
<thead>
<tr>
<th>Assembly Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use Quality Values</strong></td>
</tr>
</tbody>
</table>
The Assembly Tab contains some detailed settings related to the use of forward-reverse constraints. All these settings are disabled when Use Forward-Reverse Constraints is unchecked, and are enabled and editable when checked.

Forward & reverse reads must be named identically up to the delimiter (default = dot), and must contain paired suffixes thereafter (e.g., .s & .r or .x & .y). Suffixes may be added and deleted. Only the letter immediately following the delimiter is recognized as the suffix. Other letters following it, if any, do not contribute to the identity as forward vs reverse end sequence; however, they do distinguish one forward/reverse pair from another, thus making two sequence pairs with the matching reverse/forward end sequence. Minimum & maximum distances may also be edited.

**CAUTION** Take care when defining file names at the sequencer, especially when a project is loaded in batches over several days. Entries with misplaced suffixes (e.g., <filename>F.<projectName>, instead of <filename>.F<projectName>) will be overlooked by the constraints feature, although they may be included in an assembly.

- Min. Dist. – Minimum Distance between the forward and reverse reads. Default: 0.
- Max. Dist. – Maximum Distance between the forward and reverse reads. Default: 6000.
- Delimiter – the letter separating main part and suffix part in read names. In cases that there are multiple occurrences of the designated delimiter, the rightmost delimiter is selected.

The values of minimum and maximum distance are uniformly applied to all sequence pairs.

### Clipping Tab

Sequence clipping based on quality and similarity is an essential step in the overlap computation and assembly process and ensures validity and correctness. You can control and fine-tune the clipping process through adjusting the parameters on the Clipping Tab (Figure 30.14). Note that the clipping is irreversible and iterative within each project. When multiple assemblies are made within a project, each round of clipping applies to the sequences that had been clipped in the last round of assembly. This implies that a later attempt to assemble may affect the existing assemblies. In such cases, ContigExpress generates a warning message (Figure 30.15). If you
choose to proceed, click on the Yes button. The affected contigs will be dissolved. If you choose not to proceed, click on the No button and the current assembly will be stopped.

**Use Automatic Clipping** – Ends of sequencing reads are usually unreliable and low in QV values. CAP3 comes with a mechanism to detect and clip these poor end regions based on sequence similarities with or without QVs. Clipping is done before the computation of overlaps. Note: trimming is available in ContigExpress, in addition to CAP3 clipping, and can be carried out before loading input sequences into assembly projects. Default ON.

- **Match Score Factor** – Also used in calculating Similarity Score for computing overlaps, but only editable in the Clipping Tab. Match Score Factor is a positive integer to award each match between two bases from the pair of sequences being compared during the banded Smith-Waterman alignment. Default: 2.
- **Mismatch Score Factor** – Also used in calculating Similarity Score for computing overlaps, but only editable in the Clipping Tab. Mismatch Score Factor is a negative integer to penalize each mismatch between two bases from the pair of sequences being compared during the banded Smith-Waterman alignment. Default: -5.

- **Gap Penalty Factor** – Also used in the computation of overlaps, but only editable in the Clipping Tab. Gap Penalty Factor is a positive integer to penalize each gap extended during the banded Smith-Waterman alignment. Default: 6.

- **Quality Cutoff for Clipping** – Quality Cutoff for Clipping applies to the clipping of a poor end region for each read when QVs are provided. It is not used when QVs are not available. The specified value is used to find the low-quality ends of reads, where the quality value of a base is considered low if it is less than this value. Default: 12.

- **Clipping Range** – Clipping Range applies to the clipping of a poor end region for each read regardless of whether QVs are available. The value is used to extend the ranges for clipping further away from the ends based on the low-quality positions at each end as determined with the Quality Cutoff for Clipping value. The larger the value of Clipping Range, the more extensive the clipping for poor end regions. Default: 250.

- **Min Num Good Reads at Clip Position** – This is the depth of good coverage at the clip position. It applies to the clipping of a poor end region for each read regardless of whether QVs are available. Depending on the actual depth of coverage, the Minimum Number of Good Reads at Clip Position parameter determines the exact clipping position within the clipping region. The larger this value is, the more extensive the clipping becomes. Default: 3.

Overlap Tab

Overlaps between reads are computed immediately following sequence clipping. This step serves as the basis for contig construction. In order to ensure the quality, each overlap is evaluated with a few measures (Figure 30.16). You can change the rigor of overlap computation. The parameters are explained below.

![Assembly Setup – Overlap tab](image)

*Figure 30.16 Assembly Setup – Overlap tab*
Vector NTI Advance 10 User’s Manual

- **Overlap Length Cutoff** – Minimum length required for the length in base pair of all the overlaps. Default: 40.

- **Overlap Percent Identity Cutoff** – Minimum percent identity of the overlap. Default: 0.80.

- **Overlap Similarity Score Cutoff** – Minimum similarity score for an overlap. The score is computed as the sum of match, mismatch or gap scores for each pair of bases weighted by QVs. Default: 900.

- **Base Quality Cutoff for Difference** – Determines the minimum overlap quality by examining the differences of the overlap at bases of high quality values. This is useful only when QVs are available. The Base Quality Cutoff for Difference value defines “high quality” bases. Default: 20.

- **Max QV Sum at Difference** – Applies only when QVs are available. Each overlap is given a “difference score” based on the QVs and the Base Quality Cutoff for Difference. An overlap with a difference score over the Max QV Sum at Difference value is excluded from contig construction. Default: 200.

- **Max Clearance for Error Rate** – Applies even and especially when QVs are not available. If the error rate of the overlap is greater than the sum of those of the overlapping fragments plus this value, the overlap is not used for assembly. The smaller the value, the better the quality control of the overlaps. Default: 30.

- **Band Expansion Size** – This parameter specifies band expansion size. The program automatically determines a minimum band of diagonals for an overlapping alignment between two sequence reads. The band is then expanded, in each direction, by a number of bases specified here. This affects the computation of both potential overlaps and true overlaps. Default: 20.

- **Max Gap Length** – This is the maximum gap length allowed in an overlap. Default: 20.

- **Max Num Word Matches** – This parameter controls the fast method for finding potential overlaps between a pair of sequences. For each word in one read, at most Max Num Word Matches occurrences of the word in the other read are considered for non-gapped extension. A larger value forces the program to consider more word matches at the expense of time and computer memory. Default: 300.

- **Max Overhang Percent Length** – This parameter controls the different overhang regions before or after the aligned region. It is defined as 100 times the total length of the different overhang regions divided by the length of the overlap. Overlaps with a value greater than the maximum cutoff are not used for assembly. Default: 20.

**Contig Tab**

Contig Tab (Figure 30.17) includes parameters affecting construction of contig and multiple sequence alignments, thus the consensus sequences. QVs are used extensively during this process, if they are available. If unavailable, the program assigns a QV of 10 for each base.
Three of the parameters, Match Score Factor, Mismatch Score Factor and Gap Penalty Factor, are editable only on Contig Tab.

- **Match Score Factor** — Match Score Factor is a positive integer to award each match between the existing alignment and the sequence being added when calculating the score of global alignments. Default: 2.

- **Mismatch Score Factor** — Mismatch Score Factor is a negative integer to penalize each mismatch between the existing alignment and the sequence being added when calculating the score of global alignments. Default: -5.

- **Gap Penalty Factor** — Also Gap Penalty Factor is a positive integer to penalize each gap extended when calculating the score of global alignments between the existing alignment and the sequence being added. Default: 6.

- **Min Num Constraints Correction** — This is the minimum difference between the numbers of constraints satisfied in the current assembly and in the alternative assembly. A difference greater than this value, if the contig is also supported by an alternative set of overlaps, results in the alternative join.

- **Min. Num Constraints Linking** — This is the minimum number of constraints for reporting a link between two contigs.

**Description Tab**

Any description of the assembly can be entered here (Figure 30.18). Clicking the Default button on this tab displays a text description of all the available parameters at their default values. Any value can be edited by typing, if the corresponding parameter has been changed elsewhere in the Assembly Setup. Or you can supplement this description with information of your own. All
information entered here is viewable in the **Assembly Summary**... item on the shortcut menu of the assembly icon in Tree Pane. You can save this description as a text file.

**Figure 30.18 Assembly Setup – Description tab**

**Lite Settings Tab**

In ContigExpress, you can create two different types of contigs, Full Contigs or Lite contigs. On the **Lite Settings** tab (Figure 30.19), you can specify the type of contigs you want to assemble in the project.

In Full contigs mode, chromatogram data are retrievable and sequence editing performed in the Contig Viewer are reflected in the individual fragment files.

Lite contigs disregard fragment chromatogram data and there is no dynamic association between a Lite Contigs and its component fragments. Editing done on Lite Contigs is NOT reflected in the original fragment sequences (original sequences remain unedited). Assembly in
Lite Contig mode reduces memory consumption and is, therefore, the preferred contig type for assembling large projects.

**Figure 30.19 Assembly Setup – Lite Settings tab**

<table>
<thead>
<tr>
<th><strong>Lite Settings Parameters</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discard chromatogram data on import</strong></td>
</tr>
<tr>
<td>Check this box to discard sequence file chromatogram data when</td>
</tr>
<tr>
<td>the sequences are imported. Check this box before you import</td>
</tr>
<tr>
<td>your sequence reads. <strong>Note:</strong> you can also discard</td>
</tr>
<tr>
<td>chromatograms for all fragments after import using Project</td>
</tr>
<tr>
<td>&gt; Discard All Chromatograms command.</td>
</tr>
<tr>
<td><strong>Full Contigs only</strong></td>
</tr>
<tr>
<td>Check this box to perform assembly in Full Contig mode. You</td>
</tr>
<tr>
<td>can retrieve chromatograms and maintain dynamic links</td>
</tr>
<tr>
<td>between the contigs and their component reads.</td>
</tr>
<tr>
<td><strong>Lite Contigs only</strong></td>
</tr>
<tr>
<td>Check this box to perform assembly in Lite Contig mode. All</td>
</tr>
<tr>
<td>linkage between the contigs and sequence reads will be lost</td>
</tr>
<tr>
<td>and no chromatogram is retrievable in the Contig Viewer.</td>
</tr>
<tr>
<td><strong>Lite contigs if selected more than &lt;#&gt; fragments</strong></td>
</tr>
<tr>
<td>Creates Lite Contigs during the assembly process only if the</td>
</tr>
<tr>
<td>selected number of fragments is greater than the number</td>
</tr>
<tr>
<td>specified; otherwise creates full/regular contigs.</td>
</tr>
</tbody>
</table>

**Table 30.5 Light Settings Options**

Registering Assembly Setup Changes

Click **OK** to apply the entered parameters; they will apply to the future assemblies in the project until updated. Click **Cancel** to abort the action.
Editing Project Item Properties

To edit the name, description, or comment about a fragment or contig, right-click on the fragment or contig in the CE Project Explorer List Pane and select Edit on the shortcut menu. In the Edit dialog box (Figure 30.20) you can change the name, description and comments of the fragment:

Figure 30.20 Adding comments to fragments/contigs

Click OK to confirm the changes; all instances of the item throughout ContigExpress are updated to reflect the changes. Click Cancel to abort the action.

Viewing Project Item Properties

To view the properties of an item in a ContigExpress project, highlight the item (assembly, contig, fragment, etc.) and click the Properties button ( ) or select Edit > Properties. The corre-
The responding keystroke is ALT-ENTER. This opens the Properties dialog box (Figure 30.21) where you can review the properties of the selected item(s).

![Figure 30.21 Viewing properties of fragments/contigs](image)

**Open Contig(s)/Fragment(s)**

To view or edit a contig or fragment, select it in the Project List pane and double-click it or select **Open** in its shortcut menu.

The Contig or Fragment window for the selected item is opened.

For further information on how to work in the Workspace, Contig, and Fragment Windows see Chapter 31 and Chapter 32.

**Dismissing an Assembly**

Dismissing an assembly deletes the selected assembly from the open project. When an assembly item is selected in the Tree Pane or when the List Pane is active and shows the contents of an assembly, click the **Dismiss Assembly** button ( ) or select **Edit > Dismiss Assembly** or the comparable option on the shortcut menu. If the assembly has “child” assemblies, they will be dismissed as well. The Project Explorer display window is updated to reflect the changes.

**Dissolving a Contig**

Dissolving a contig removes the selected contig from the assembly, but leaves the rest of the assembly intact. To execute the Dissolve Contig command, select the contig in the List or Tree Pane of the Project Window and choose **Dissolve Contig from Project menu** or from the shortcut menu. Once the contig has been dissolved, the Project Explorer display window is updated to reflect the change.
Deleting Contig(s)/Fragment(s) from a Project
To delete selected item(s) from the project, select the item in the List Pane and click the Delete from Project button ( ), press the keyboard DELETE key, select Edit > Delete from Project or Delete from Project from the shortcut menu. If any of the items to be deleted are included as a fragment in any assembly, you will be warned that all assemblies (and their child assemblies) that contain the item will be dismissed. You can confirm or cancel the action.

Renaming an Item
To rename a project item, select the item and click the Rename button ( ), select Edit > Rename or single click inside the name of the item to be renamed. Enter the new name in the enabled text box. If the item that has been renamed is used elsewhere in the project, all instances of the item are renamed.

Operations Changing Project Items
Converting a Contig to a Fragment
A fragment with the same sequence as a contig’s consensus and the same name as the original contig can be created and added to the project. To convert an existing contig to a fragment, select File > Convert Contig to Fragment on the menu bar or select Convert Contig to Fragment on the shortcut menu.

Making a Reverse Complement
To generate reverse complement copies of selected contigs and fragments from the Project Explorer List Pane, select the item and choose Make reverse complement from the shortcut menu. If a contig is selected for this action, it is internally converted to a fragment and then the action will be applied to it.

The reverse complement copies of the selected items are added to the project with names Copy of <name >, derived from the names of the original items. The new item appears in the List Pane.

Finding Sequences
There are two find features available in the ContigExpress Project Explorer window. The Find Fragment feature allows you to search for and find the location of fragments in the project according to fragment name (See Finding Fragments on page 440.). The Find Sequence feature allows you to find fragments and contigs in the project that contain the sequence that you specify.

Launch the Find Sequence dialog box by choosing Edit > Find Sequence from the Project Explorer menu or by pressing CTRL+ SHIFT+ F on the keyboard. The Find Sequence dialog box displays (Figure 30.22):

Figure 30.22 Find Sequence dialog
Type the sequence you want to search for in the Sequence box and set the search parameters:

- **Revert** – Reverse complements the sequence entered in the Sequence box.
- **Search Options** – Allows you to search fragments or contigs only.
- **Search area** – Specify the area of the project you want to search: the assembly currently highlighted only, from the current assembly forward, or the entire project.

Once you have entered the sequence and set the search conditions appropriately, click the **Find** button to begin the search. Search results display in the lower portion of the Find Sequence dialog box (Figure 30.23):

![Find Sequence dialog](image)

*Figure 30.23 Search results in Find Sequence dialog*

Double-click on an fragment/contig name in the results section of the Find Sequence dialog box to highlight the fragment/contig in its location in the Project Explorer List Pane.

**Blasting Ultimate ORF Collection**

Ultimate ORF Blast Invitrogen has generated a database of proprietary ORFs, the Invitrogen Ultimate™ ORF collection. With the Ultimate ORF BLAST tool and using a fragment selected in Contig Express Contig Viewer as the query sequence, you can launch a search against the proprietary ORF collection.

To launch Ultimate ORF Blast, select **Tools > Ultimate ORF Blast**. Choose the **Range** (Whole, Selection Only, or Current CDS feature) and **Sequence display** (Single or Double Stranded). Click OK. See **Ultimate ORF Blast** on page 467.
**Generating an Assembly Summary**

To write a text file with an assembly summary, select an assembly in the Tree Pane, and select **Assembly Summary...** on the shortcut menu. This writes an “Assembly Summary” file with a “.log” suffix, summarizing the contig(s) by consensus length, sequence, number of fragments and date of generation. In addition, summaries of each fragment in a contig appear with name and position. Finally, data recorded in the Description Tab in the Assembly Setup dialog is summarized. Read the .log file in NotePad.

In the Write Assembly Summary dialog box, enter a file name for the summary and click Save. If the file with this name already exists, you can overwrite the file. Click Cancel to abort the action.

**Trimming Fragments**

Trimming fragments is often performed to remove unreliable (ambiguous) residues. To trim one or more fragments selected in the Project Explorer List Pane, select **Edit > Trim Selected Fragments <trim type>**. A Fragment Trimmer dialog box opens, listing all of the selected fragments and/or suggested operations, depending on which type of trimming operation you’ve chosen (see Chapter 32, page 493). Click **OK** to proceed with trimming the fragment(s). If no action is suggested, the **OK** button is disabled.

Notes on Fragment Trimming:

- This group of actions is used to trim the selected fragments according to various criteria.
- Because “trimming” applies only to fragments, if contigs are selected, they will be skipped.
- If a fragment selected for trimming is currently open in its Fragment Window, initiating the trimming mode in Project Explorer automatically forces the fragment into the read-only mode in the Fragment Window to avoid possible inconsistencies in the project after a successful trimming.
- If any of these fragments were already changed in their respective windows, select **Yes** to save the changes or **No** to lose the changes or **Cancel** to skip the fragment in question.
- Project Explorer reflects the changes if fragments are trimmed. If any of the changed fragments are open in Fragment Windows, the respective windows are updated as well. For further information on the Fragment Trimmer dialog boxes, including modifying trimming settings, refer to Chapter 32, page 493 of this manual.

**Calling Secondary Peaks for Fragments**

If conflicting residues appear on a chromatogram, you can alter the nucleotides as you wish. With one or more fragments selected in the Project Explorer List Pane, select **Edit > Call Secondary Peaks for Selected Fragments**. The Call Secondary Peaks dialog box opens (see Chapter 32, page 502), listing all selected fragments will be presented to the user. Click **OK** to continue.

Notes on Calling Secondary Peaks:

- Because this operation applies only to fragments, if contigs are selected, they will be skipped.
- If a fragment selected for secondary peak search is currently open in its Fragment Window, initiating this operation in Project Explorer automatically forces the fragment into the read-only mode in the Fragment Window to avoid possible inconsistencies in the project after a successful search.
- If any of these fragments were already changed in their respective windows, select **Yes** to save the changes or **No** to lose the changes or **Cancel** to skip the fragment in question.
For further information on the Call Secondary Peaks dialog box, refer to Chapter 32, page 502 of this manual.

Other Operations in Project Explorer

Project Explorer Viewing Options

To review or modify viewing options in Project Explorer, highlight a fragment or contig in the List Pane and select View > Options from the menu. This opens the Options dialog box, which has several tabs.

Note: Option settings for contigs and fragments are independent of each other. If a fragment is highlighted when the Options dialog box is opened, display options for fragments are visible in the dialog box; if a contig is highlighted when the Options dialog box is opened, display options for contigs are visible.

Details View Tab

The Options dialog box initially opens to the Details View tab (Figure 30.24).

The Details View tab controls the columns displayed in the List Pane Details View mode and related functions as follows:

- **Available Fields** – Shows the list of fields that you can add to the List Pane display. Move columns from the Available Fields box to the Displayed Fields box (and vice-versa) by highlighting the desired column name and clicking the >> or << button.

- **Displayed Fields** – Shows the fields that are currently being shown in the List Pane display. The top-to-bottom order of columns in the Displayed Fields list dictates the left-to-right column order shown in the List Pane; top to bottom order in the Displayed Fields list corresponds to left to right order in the Project Explorer List Pane.

- **Order** – Allows you to rearrange the order of the fields in the Displayed Fields list. To change the order, select a field and click the Move Up or Move Down button.

- **Alignment** – Controls the alignment of text in the column headers in the List Pane. To alter the text alignment for a column, highlight the column name in the Displayed Fields list and select the appropriate radio button.

Figure 30.24 Options for displaying column headers in List Pane
- **Minimum number of overlapping fragments to calc unambiguity** – Allows you to specify the minimum number of fragments used for calculating unambiguity for a contig. Percent unambiguity is shown in the **Unambiguousness** field.
- **Save the last viewed assembly** – Saves the display format of the last viewed assembly.

**Status Bar Tab**

The second of the Option tabs is the **Status Bar** tab (Figure 30.25).

![Options screenshot](image)

*Figure 30.25 Options for display in Explorer Status Bar when items are selected in List Pane*

- The **Status Bar** tab allows you to choose the type of information displayed in the status bar when a single item is selected in the List Pane:
  - **1 Item Selected** – Select this option to show the text “1 item selected” in the status bar when an item is selected in the List Pane.
  - **The Value of the Item’s Field** – Select this option and a field in the list to show the value of that field for the item selected in the List Pane.
Copy Fragment/Contig Tab

The third of the Option tabs is the Copy Fragment/Contig tab (Figure 30.26):

The Options for export file type tab (Figure 30.26) allows you to define the file format used when you drag an item outside of the Project Explorer or copy it to the Windows Clipboard.

Copying Selected Fragments or Contigs

In Project Explorer, to copy a fragment or a contig select it and click the Copy button ( ).

Click the Paste button to paste the copy in the List Pane ( ). The copied object is added to the list of objects and named “Copy of <selected object>”.

Printing

You can print the contents of the Tree Pane in the CE Project Explorer window (Figure 30.27). To review what the printout of the Tree pane looks like before it is printed, select Project > Print Preview:

In the Print Preview dialog box, you can resize the displayed output, send it to the printer, or press Close to return to the Project Explorer. To define your printer options, select Project > Print Setup. A standard Window Printer Setup dialog box opens.
To print a copy of the Tree Pane data, activate the Tree Pane. The Tree pane is always printed in the History view mode with the tree completely expanded. Select Project > Print. In the Print dialog box, click OK to print or Cancel to abort the action.

Exporting Items from the Project

To export contigs or fragments selected in the Project Explorer List Pane, choose Project > Export Item > To <file> from the Project Explorer menu or select Export Contig/Fragment > To <file> on the contig/fragment shortcut menu. ContigExpress can export items in the following formats: GenBank (*.gb), FASTA (*.txt), EMBL (*.txt), and plain text sequence file (*.txt). Choose Project > Export Item > Send to VectorNTI, and the selection will open directly in a VectorNTI viewer.

In the Export To dialog box, select or enter a file name and destination for the exported item. Click Save to export the item to the entered file. You can overwrite an existing file or click Cancel to abort the action.

Alternatively, you can select one or several contigs or fragments in the List pane and drag and drop them to Windows Explorer or Vector NTI Explorer. The export format for dragged and dropped files is defined by the setting on the Copy Fragment/Contig tab of the Options dialog box (see page 459).

You can return to other Vector NTI Advance programs or send ContigExpress data to AlignX by selecting Align > <AlignX options> and to BioAnnotator by selecting Analyses > BioAnnotator > Analyze Selected Molecule.

Saving and Closing a Project

To Save a Project to disk, click the Save button ( ), select Project > Save, or use the keystroke CTRL+S. Enter a file name and destination in the Save As dialog box. After the project is successfully saved, the Save toolbar button and menu item are disabled until you make any changes in the project.

To close a Project, select Project > Close Project. When you select this action, the Project Explorer first closes the current project, giving you the option of saving any changes. To save any changes, click Yes in the dialog box that opens. If you select No, all changes are lost.

Exit Project Explorer

To exit Project Explorer, click on the Explorer close box, select Project > Exit, or use the keyboard shortcut, Alt-F4. This operation closes the ContigExpress application.
Introduction

ContigExpress offers two workspaces where you can work with fragments or contigs. Fragment Viewer allows you to work with individual fragments. Contig Viewer allows you to work with contigs or the fragments comprising them. To open fragments or contigs in their respective viewers, select the item in Project Explorer and double-click on it or select Open from the shortcut menu. Depending upon the item from which it was launched, a Fragment Viewer or a Contig Viewer opens. When projects are assembled in Full Contig mode, Fragment Viewers and Contig Viewers may be opened together in the same workspace; the Fragment Viewer, however, is in Read-Only mode. This section reviews elements common to both types of viewers.

Features of the ContigExpress Workspace

These workspace windows have many of the same features as the other applications of Vector NTI Advance: a menu bar, toolbars, and three panes and a status bar.

As you are introduced to various operations in this chapter, the toolbar buttons are displayed as needed. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.
The windows are organized and managed like any Windows application. Several windows can be opened in one workspace (Figure 31.1).

Fragment and Contig Viewers are divided into three panes. Split bars allow re-sizing the three panes. Scroll bars appear when the contents of a pane exceed the viewing area, facilitating easier viewing of pane contents.

If you prefer, you may view contigs or fragments in a different layout. From the View menu, go to Edit Pane Layout and toggle between the standard, tabbed or custom layout options. To design a custom layout, right click on entries in the Pane Manager, design a layout, then drag and drop the desired pane(s) to the chosen layout. The layout you select will be the default one until a new layout is chosen.

Switching Between Panes

As in the other Vector NTI Advance applications, to apply a command from the Main Menu in a given pane that pane must be active. The active pane can be toggled with the Panes buttons on the Pane Selection Toolbar or by clicking anywhere in the pane you wish to activate. The keyboard strokes are F6 (cycling clockwise through the panes) and SHIFT + F6 (cycling counterclockwise).

Status Bar

The Status bar shows menu item help and information about the cursor and selection positions in the active window (Figure 31.2).

The menu item help on the Status Bar is a short description of a selected menu or toolbar item.
The Position Box shows the current cursor position if relevant. You can position the cursor by clicking on this box. In the Fragment Viewer, the display is in the form original position/position post-trimming. In the Contig Viewer, the Position Box can apply to fragments or the consensus. On fragments, the display indicates position in the Contig (gaps included)/original position in the fragment. On the consensus, the display indicates position in the Contig (gaps included).

The Selection Box displays the start and end coordinates of selections in the active window. In the Fragment Viewer, each coordinate is in the form original START position – START position post-trimming : original END position - END position post-trimming. You can also position a selection by clicking on this box and checking Do Not Count Gaps on/off as desired. In the Contig Viewer, the Selection Box can refer to fragments or the consensus, with the guidelines described above. Click this box to set a selection on a fragment (using fragment or contig coordinates) or on a consensus (using Contig coordinates).

General User Actions

Finding a Feature in Other Panes

A feature selected in a Feature Table folder in the Text Pane can be concurrently selected in both other panes. Select the feature, then click the Find button ( ) on the Pane Toolbar or select Find on the shortcut menu opened from a folder.

Note: Because some features are circular, they bypass the endpoints of the fragment. Since ContigExpress does not support circular selections, such a feature will not be correctly located. The program warns you in such cases.

Editing Fragment or Contig Properties

To edit some of a selected contig or fragment’s properties from the Text Pane, double-click on the first line with the fragment or contig name or select Edit from the associated shortcut menu.

In the Edit dialog box (Figure 31.3), change the name, description or comments:

![Figure 31.3 Edit Properties dialog](image)

Click OK to confirm the changes or Cancel to abort the action.

After selecting OK, every instance of the fragment throughout ContigExpress is updated to reflect the changes.
Saving a Current Window

To save a changed item/apply changes made in a Fragment or Contig Viewer to the project, select **Apply** from the Fragment or Contig Menu, use keystroke **CTRL-S**, or click the **Apply** button ( ).

If a change affects other project items, Project Explorer makes the necessary changes, and updates its display and the currently opened windows, if needed.

Closing a Current Window

To close the current window, select **Close** from the Fragment or Contig menu, click on the respective Close Box, or use keystroke **CTRL + F4**. If the item in the window you are closing has not been saved since its last change, you can confirm saving the changes in the dialog box that opens.

If the item in the window you are closing has not been saved since its last change, you can confirm saving the changes in the dialog box that opens.

Closing All Windows

To close all open windows (thus closing the Workspace), select **Close All** from the Fragment or Contig menu, click on the workspace Close Box, or use keystroke **ALT + F4**.

Selecting All Data

To select all data in the current pane, select **Edit > Select All** or choose **Select All** on the shortcut menu.

Deleting Data

To delete a data selection from an item in the active pane, select **Edit > Delete Sequence From <selected bp range>** or press the **DELETE** key on the keyboard. If no selection is currently defined, one nucleotide is deleted at the cursor position with this action. (If you delete in error, click on the **Undo** button ( ).

The following additional actions can be performed in ContigExpress windows. Details are given in Chapter 3.

- Resizing the panes with the split bars.
- Scrolling data inside the panes using scroll bars and arrow keys
- Opening and closing Text Pane folders
- Opening shortcut menus
- Copying selected molecule and fragments
- Copying screen and pane contents
- Pasting files and text
- Exporting data to other Vector NTI applications or external programs
- Printing from ContigExpress windows

External Tools Usage

ContigExpress is integrated with the other components of Vector NTI Advance, making it easy to launch Vector NTI, AlignX and BioAnnotator from ContigExpress and to easily exchange data with them. It can also easily export data to third-party tools available on the Internet.
To invoke the tools, select **Analyses > Web Analyses > <tool category> > <tool>, Align > Web Alignment Tools > <tool category> > <tool>, or Tools > Online Resources > <tool>**. The tools listed in these menus will work with the currently selected items.

For more information on the External Tools, refer to other chapters in the Vector NTI documentation.

**Fragment Viewer**

**Launching a Fragment Viewer**

To open a Fragment Viewer, select a fragment in CE Project Explorer and double-click on it or select **Open** on its context-sensitive shortcut menu. Alternatively, Fragment Viewers may be launched from the respective mini-windows in the Fragment Trimming dialogs (Chapter 32). Finally, when a contig is assembled in Full Contig mode, one or more instances of the Fragment Viewer may be opened in Read-Only mode by double-clicking on the respective fragments in any pane of the Contig Viewer. Note that if an assembly was performed in Lite Contig mode, it will not be possible to open Fragment Viewers from the Project Explorer or the Contig Viewer.

**Features of a Fragment Viewer**

As in the other applications of Vector NTI Advance, the ContigExpress Fragment Viewer (Figure 31.4) consists of a menu bar, two toolbars and is divided into three panes: Text Pane, Sequence Pane, and Chromatogram Pane. As described above (**Features of the ContigExpress Workspace**), the panes can be organized in different layouts. A shortcut menu associated with specific folders or objects can be opened with a right click while the cursor is on the object. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

![Image: Fragment Viewer - standard layout](image)

As in the other Vector NTI Advance applications, to apply a command from the Main Menu in a given pane, that pane must be active. The active pane can be toggled with the **Switch Pane** buttons ( ) in the Pane Selection Toolbar or by clicking in the pane you want to activate. You can also activate the panes with the F6 and SHIFT-F6 keyboard buttons.
**Text Pane**

The Fragment Viewer Text Pane contains a thorough description of the fragment, including length, date of creation, and chromatogram data.

Text Pane manipulations are outlined in the previous section and are also summarized in Chapter 3.

The Text Pane data is divided into a hierarchy of folders and subfolders. A Text Pane can contain the following folders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description</td>
<td>Molecule type, form (for DNA only), length, etc.</td>
</tr>
<tr>
<td>Proprietary fields</td>
<td>User defined fields (for molecules which came from Vector NTI)</td>
</tr>
<tr>
<td>Standard fields</td>
<td>GenBank/SWISS-PROT-like fields: keywords, division, original accession numbers, etc</td>
</tr>
<tr>
<td>Comment</td>
<td>Arbitrary text of any length associated with the molecule</td>
</tr>
<tr>
<td>References</td>
<td>Bibliographic references (in GenBank/SWISS-PROT format)</td>
</tr>
<tr>
<td>Feature Map</td>
<td>List of molecule features</td>
</tr>
<tr>
<td>Chromatogram data (if available)</td>
<td>Data imported from fragment file processed by sequencing machine (ABI files, for example)</td>
</tr>
</tbody>
</table>

*Table 31.1 Fragment Viewer Text Pane folders*

**Sequence Pane**

The Sequence Pane shows the nucleotide sequence of the fragment with feature signals. In this pane, the nucleotide sequence can be viewed, edited and manipulated and the sequence display can be formatted. Any sequence changes are immediately reflected in the Chromatogram Pane. Regions of the sequence can be copied and pasted to/from the Clipboard. For more details, refer to Chapter 3.

In addition to the sequence, the Sequence Pane displays a molecule’s nucleotide sequence features. Features are shown with a horizontal wavy line above the sequence (direct strand features) or below the sequence (complementary strand features) together with the feature label.

Navigating the Sequence Pane Using the Keyboard

Pause the text cursor over a position in the sequence for a few seconds to display a tooltip defining the exact position of the cursor. The cursor position or selection position is always indicated on the status bar following a left mouse-button click or selection.

Standard keyboard keys may be used to navigate the sequence and position the cursor:

<table>
<thead>
<tr>
<th>Key</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEFT</td>
<td>Move cursor to previous position</td>
</tr>
<tr>
<td>RIGHT</td>
<td>Move cursor to next position</td>
</tr>
<tr>
<td>UP</td>
<td>Move cursor position one line up</td>
</tr>
<tr>
<td>DOWN</td>
<td>Move cursor position one line down</td>
</tr>
</tbody>
</table>

*Table 31.2 Navigating the Sequence Pane using the keyboard*
Another way to define an exact cursor position is to click the Set cursor Position box on the status bar. In the Set cursor Position dialog box that opens, enter the cursor position and click OK.

Finding Sequence Fragments

To search for a specific sequence in the Sequence Pane, activate the pane, then click the Find button ( ) on the toolbar, or select Edit > Find.

![Find Sequence dialog](Figure 31.5 Find Sequence dialog)

In the Find dialog box (Figure 31.5), enter the sequence in the text box, with criteria for the search, such as the strand to search, (direct or complementary), allow ambiguous symbols, and the search direction from the current cursor position. If the matching sequence is found, ContigExpress selects it and positions both the Sequence Pane and Chromatogram Pane to make it visible. If the sequence is not found, you are informed and the cursor remains in its current position.

Open Reading Frames

To display ORFs in the sequence, activate the Sequence Pane and choose View > Show ORFs from the menu or press the Show ORFs button ( ) on the toolbar. If the sequence display is double-stranded, ORFs for both strands display; if the sequence is single-stranded, ORFs for the direct strand only display. ORF definition can be adjusted on the View Options > ORF tab, accessed from the View menu or Toolbar button ( ). For details on displaying ORFs in the Sequence Pane, see page 213.

Searching for Ambiguous Symbols

To search for ambiguous symbols quickly, click the Find Previous Ambiguous ( ) or Find Next Ambiguous ( ) buttons from the View menu or the respective buttons on the Toolbar.

Ultimate ORF Blast

Invitrogen has generated a database of proprietary ORFs, the Invitrogen Ultimate™ ORF collection. With the Ultimate ORF BLAST web tool in Vector NTI Advance, you can BLAST your frag-
ments against the Ultimate™ ORF collection. Using this tool, you may determine that your sequence has been previously characterized and the clone containing it is commercially available from Invitrogen.

To launch Ultimate ORF Blast, select the all or part of a fragment, and select **Tools > Ultimate ORF Blast**. The Sequence Data dialog box opens, allowing you to set the search parameters:

- **Range** selects the range of the target sequence: the entire sequence, a selected sequence or a selected CDS.
- **Strand** selects either the direct or complementary strand. Click **OK** to begin the search or **Cancel** to abort the search process.

The Ultimate ORF Blast tool copies your target sequence and pastes it into the Ultimate ORF Browser Sequence Text entry box located on the Invitrogen webpage (Figure 31.6), which opens automatically when you launch this tool.

![Image of Ultimate ORF Browser](image)

**Figure 31.6 Ultimate ORF Browser**

Select the species to search the sequence query against from the checkbox located above the Sequence Text entry box. To proceed with your search, click the **BLAST** button. The Blast dialog box appears, informing you that it is blasting your sequence.

Once your search is completed, the Ultimate™ ORF Browser displays your search results. If your search identifies similar ORFs, the target sequence, or Ultimate™ ORF clone, is commercially available through the Invitrogen website, http://www.invitrogen.com.

**Miscellaneous Operations in the Sequence Pane**

For additional operations in the Sequence Pane, refer to Chapter 3:

- **Changing Sequence Pane Properties** By default, the Sequence Pane displays the direct strand only. **Show two strands** display option is available from the **View** menu or the Toolbar button ( ).
- **Translating sequences** Any selection of more than 3 base may be translated in the reading frame defined by the selection, using the commands from the **View** menu or the Toolbar buttons ( ). Translation may appear in 3- or 1-letter code, and may be
from the direct or complementary strand. Clear All Translations is also available from the menu or Toolbar.

- Changing text display attributes The Grouping command from the View menu allows you to display bases in groups of 1, 3, 5, 10 and 20. You can also Set Number of Blocks per line. These options are duplicated on the View Options > General tab.

**Chromatogram Pane**

The Chromatogram Pane displays (Figure 31.7) displays the chromatogram of the fragment if it was initially provided in the original imported file used to create the fragment. A chromatogram track shows:

- a non-editable, color-coded sequence track, with coordinates
- a duplicate sequence at the top of the pane where that can be edited
- a vertical slide control for adjusting the dynamic range of the peak heights

![Figure 31.7 Details of fragment Chromatogram Pane](image)

Any editing changes in the Chromatogram Pane are immediately reflected in the Sequence Pane and vice versa. Regions of the sequence may be copied and pasted to/from the Clipboard.

**Showing/Hiding Traces**

By default, all traces are shown in the Chromatogram Pane of the Fragment Viewer. Traces for each base can be toggled off and on independently in the Chromatogram Pane. Choose View > Show Traces > <base> Trace from the menu or click the Show Trace button for the particular base on the toolbar to hide (or show) the trace for that base.

**Navigating the Chromatogram Pane Using the Keyboard**

Standard keyboard arrow keys can be used to navigate the chromatogram.

A way to define an exact cursor position is to pause the cursor over a residue for a moment allowing a popup label to identify its position. The tooltip displays the peak height for all 4 traces. In addition, if the fragment has associated quality values, they will also be displayed.

To position the cursor specifically, press the Set cursor Position box on the status bar. In the Set cursor Position dialog box that opens, enter the cursor position and click OK.
Resizing the Chromatogram

Chromatograms may be resized to facilitate examination of peak shoulders. To re-size the chromatogram, use the Zoom In (🧬) or Zoom Out (🧬) buttons or select View > Zoom In or Zoom Out or the corresponding commands on the shortcut menu. Chromatogram peak height can be adjusted with the vertical slider control on the left side of the Chromatogram Pane.

Fragment Viewer Operations

**Editing a Sequence**

If the fragment you are editing in the Fragment Viewer is not included in any assembly, then you can edit that fragment freely. When you save your changes using the Fragment > Save As command or by pressing CTRL-S, the fragment is not saved on a disk. Instead, the fragment in the main fragments list in the current project is updated. To save your changes on disk, apply your changes first in the Fragment Viewer, and then either save the current project or export the fragment to disk. You can perform these operations in Project Explorer Window. See page 460 for more information on how to save projects and export fragments.

If the fragment you are viewing is included in a contig in this or another assembly, editing is disabled. If the fragment is in an assembly, but not part of any contig, editing is enabled.

You can edit the sequence in the Sequence or Chromatogram Panes in the same way you would edit text in a word processor with the following exceptions:

- In the Sequence Pane, only the direct strand can be edited. The complementary strand is automatically updated.
- In the Chromatogram Pane, only the top sequence can be edited. The original fragment sequence at the bottom is included for reference only.
- In both Chromatogram Pane and Sequence Pane, you can only replace one nucleotide at a time.

The valid symbols you can enter are A, T, G, C, N and other acceptable ambiguous nucleotide designations specified by IUB codes (see Appendix C). Editing changes in either pane are reflected in the other pane immediately.

Sequence editing techniques are detailed in the following table:

<table>
<thead>
<tr>
<th>Editing Operation</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete a nucleotide</td>
<td>Position the cursor to the left of the symbol to be deleted and press <strong>DELETE</strong>. (If you are editing in the chromatogram, you may also place the cursor to the right and press <strong>BACKSPACE</strong>.) In the Sequence Pane, the base drops a row and an arrow appears in its position. In the Chromatogram Pane, the position in the top sequence of the chromatogram becomes a gap.</td>
</tr>
<tr>
<td>Insert a symbol</td>
<td>Select the sequence region to be cut. Press <strong>CTRL-X</strong> or choose <strong>Edit &gt; Cut</strong>. The selected symbols are deleted from the sequence (as above) and copied to the clipboard.</td>
</tr>
<tr>
<td>Replace a symbol</td>
<td>Position the cursor where the new symbol is to be inserted in either Pane. Type a valid symbol. The new symbol is inserted and a gap appears in the chromatogram (if present).</td>
</tr>
<tr>
<td>Paste a sequence from the clipboard</td>
<td>Position the cursor where the symbols are to be inserted, or select the part of the sequence you want to be replaced. Press <strong>CTRL-V</strong> or <strong>Edit &gt; Paste</strong>…</td>
</tr>
</tbody>
</table>

*Table 31.3 Sequence editing techniques*
Cut, Copy and Paste options are also available on the shortcut menu opened from the Sequence Pane.

**Calling Secondary Peaks**

It is desirable at times to review secondary peaks in a chromatogram and change base calls. The Edit menu command *Call Secondary Peaks* opens a dialog box to search for secondary peaks that are at least some percentage of the highest peak at a base call position. Setting a value and clicking OK replaces symbols in the fragment (Figure 31.8) as described above and reports the number that were replaced. For further details on calling secondary peaks, refer to Chapter 32.

![Figure 31.8 Secondary peaks called at 80% level](image)

**Showing Deleted Peaks**

By default, peaks for deleted bases are visible in the Chromatogram Pane. Peaks for deleted bases can be removed by toggling View > Show Deleted Peaks from the menu or by clicking off the Show Deleted Peaks toggle button on the toolbar. This action informs the user that the fragment is entering Read-Only mode; viewing the fragment without deleted peaks does not alter the sequence.

When deleted bases are showing, translations proceed through the deletions without considering them. With the deleted peaks removed, the fragment can be translated across a deletion in a new reading frame and the validity of the deletion assessed. For more information regarding Showing Deleted Peaks, see Chapter 17.

**Revert to Experimental**

The original sequence of a fragment may be restored after cumulative edits resulting from multiple operations, including edits performed in the Viewer, and Ends or Vector Contamination Trim-
ming performed in the Explorer. In the Fragment menu, choose **Revert to Experimental**. This option is disabled for fragments already in a contig.

**Window Viewing Options**

The following section summarizes display options available for the Fragment Viewer. Many of the default options have been described already in context. Others have been alluded to briefly or not at all.

To modify viewing options for the Fragment Viewer, click on the **Viewing Options** button ( ) on the Toolbar or select **View > View Options**. This opens the Fragment Viewing Options dialog box (Figure 31.9):

**Figure 31.9 View Options dialog – General tab**

On the **General** tab, you can do the following:

- **Nucleotide Sequence**: change the length of a single line of a sequence, the number of blocks per line, single/double strand display in the Sequence Pane
- **Search**: amend the default settings for the Find dialog box
- **Translation**: select the use of three- or one-letter protein codes displayed in the sequence translation.
- **Gaps**: settings for the Do Not Count Gaps checkbox in the Set Cursor Position and Set Selection dialog boxes. The second checkbox, Gap Chromatogram on Base Insertion (default on) controls how the chromatogram responds when new bases are inserted into an editable sequence. You may either choose to gap a chromatogram upon base insertion or insert a new base without a gap appearing in the chromatogram (deselect).

On the **ORF** tab, you can set the parameters for ORFs display in the Fragment Viewer (Figure 31.10):

**Figure 31.10 View Options dialog – ORF tab**

- **Nested ORFs**: select if nested ORFs are shown
- **Include stop codons in ORF**: include stop codons in the ORF
- **Minimum size**: specify the minimum size of an ORF
- **Start codons**: specify the start codons
- **Stop codons**: specify the stop codons
- **Default**: set default values for ORF display
ContigExpress: Fragment and Contig Viewers Chapter 31

- **Nested ORFs**: Check the box to display ORFs that occur within the main ORF.
- **Minimum Size**: Specify the minimum ORF size in codons that you want to display.
- **Start Codons**: Define acceptable start codons.
- **Stop Codons**: Define acceptable stop codons.
- **Defaults**: Restores ORF dialog box default values.

**Contig Viewer**

**Launching a Contig Viewer**

To open a Contig Viewer, double-click on a contig in CE Project Explorer, or select **Open** on a shortcut menu opened from a contig in Project Explorer.

**Features of a Contig Viewer**

The ContigExpress Contig Viewer (Figure 31.11) consists of a menu bar, two toolbars and is divided into three panes: a Text Pane, a Graphics Pane and a Contig Alignment Pane. As described above (Features of the ContigExpress Workspace), the panes can be organized in different layouts.

A shortcut menu associated with specific folders or objects can be opened with a right click while the cursor is on the object.

**Note:** Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

![Figure 31.11 Contig Viewer - Standard Layout](image)

As in the other Vector NTI Advance applications, to apply a command from the Main Menu in a given pane, that pane must be active. The active pane can be toggled with the **Switch Pane** buttons ( , , ) in the Pane SelectionToolbar or by clicking in the pane you want to activate. You can also activate the panes with the F6 and SHIFT-F6 keyboard buttons.
Find a Fragment
To rapidly locate a fragment in a Contig, the Find Fragment search dialog can be activated from the Edit menu on all three panes. Enter a fragment name in the text window and click the Find button ( ) as shown in Figure 31.12. Asterisks (*) may be used as wild cards in the search in accord with the specific syntax of the file names (e.g. *.*_1124*). Double-clicking on a Search Result will highlight the fragment in all 3 panes.

![Find Fragment dialog](image)

Figure 31.12 Find Fragment dialog Wildcards (*) are used in accord with syntax of project file names

Text Pane
The Contig Viewer Text Pane contains a thorough description of the contig displayed in the window.

Text Pane manipulations are outlined in the previous chapter and are also summarized in Chapter 3.

The Text Pane data is divided into a hierarchy of folders and subfolders. A Contig Text Pane can contain the following folders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description</td>
<td>Name, creation date, length, etc.</td>
</tr>
<tr>
<td>Comment</td>
<td>Arbitrary text of any length associated with the contig</td>
</tr>
<tr>
<td>Fragments</td>
<td>List of fragments contained in the contig</td>
</tr>
</tbody>
</table>

Table 31.4 Contig Text Pane folders

Graphics Pane
The Contig Viewer Graphics Pane (Figure 31.13) contains horizontal arrows representing the relative positions of the fragments forming the contig. The arrowheads indicate whether the respective fragment is in the direct or complementary strand, with the names of the fragments displayed above the fragment lines. When either of the Graphics or Alignment Panes is active, one of the fragments is always highlighted with another color (usually dark blue). This highlight is coordinated with highlights in the Sequence Pane to allow you to locate the same fragment in all panes.
When the Contig Viewer first opens, all data in the Graph pane is in the Fit to screen mode. In this mode all the graphical data for the complete contig is always shown in the pane no matter how you resize the pane.

If you want to examine a part of the graph pane, you can resize the pane using the Zoom and **Fit to Screen** buttons. Resizing of the pane does not affect the current scale of the graph. For more information on using the Zoom buttons, refer to Chapter 3.
By default, zooming is done horizontally. If you want to zoom vertically, press \texttt{CTRL + Zoom In} button ( ).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure31_13.png}
\caption{Details of contig Graphics Pane}
\end{figure}
Contig Coverage in the Graphics Pane

You can view a graphical representation of contig coverage in the Graphics Pane. Display of the contig coverage graphic in the Graphics Pane view (Figure 31.14), can be turned off/on by

**Show Contig Coverage** in the View menu or the toolbar button ( ).

![Figure 31.14 Contig coverage bar- coverage category summarized in tooltip](image)

The Contig Coverage graphic appears as a bar, which spans the length of the contig and contains segments of varying pattern/color that represent the amount and type of fragment coverage in that segment. If you hover the cursor over a segment, a tooltip appears describing the coverage type/amount and segment coordinates. Types of coverage include the following:

- **Single Fragment** – area is covered by one fragment only.
- **Multiple Fragments Same Direction** – area is covered by two or more fragments that run in the same direction.
- **Both Directions** – area is covered by at least two fragments that run in opposite directions.
- **Multiple Fragments Both Directions** – area is covered by more than two pairs of fragments that run in both directions.

**Weight Graph**

Below the Contig Coverage bar is a scale or Weight Graph, for this contig. The Weight Graph gives an overview of assembly quality. You may note small green vertical lines dropping below the horizontal . If you position your cursor at one of those sites and click the **Zoom In** button repeatedly, you can see the exact nucleotide placement of each line in the graph. You will also note the negative peak in the Weight graph line, indicating a position where the assembly is
poor (Figure 31.15). Refer to the same position(s) in the chromatogram (if available) and you will see poorly resolved peaks.

In the Weight Graph, each combination of two possible sequence symbols is assigned a certain weight. By default, the highest weights are given to combinations of equal symbols, for example, “AA” or “CC”. The weights can be changed in the Viewing Options dialog of the Contig Viewer. At each position in the contig, the weight is calculated from all symbols in the fragments available for this position and from the consensus symbol. The resulting weight is plotted as a point in the Weight Graph. Dips in the line isolate places where the assembly is poor or nucleotides are mismatched.

Showing ORFs in the Graphics Pane

ORFs can be shown for the fragments in the Graphics Pane. Choose Show ORFs from the View menu or click the Show ORFs button on the Toolbar to add an ORFs display to the Graph-
ics Pane. Parameters defining an ORF may be altered on the ORF tab on the Viewing Options dialog (accessed from the View menu or the Viewing Options toolbar button).

Figure 31.16 ORF analysis of contig generated from fragments of HS3454_C02

In Figure 31.16, the three bars between the Contig Coverage and Weight Graphs represent the 3 forward reading frames of the contig, and the green segments indicate the location of their respective ORFs. Hovering the cursor on a green segment displays a tooltip with the ORF coordinates. Clicking on it highlights that ORF in both the Graphics and Alignment Panes.

Open a Fragment In Another Window

Open fragment(s) in a Fragment Viewer by double-clicking on the fragment line in the Graphics Pane, or the fragment name in the Alignment Pane.

Camera Specifics

While you can use the Camera feature to copy pane contents, only the currently visible portion is copied to the clipboard or file. In this pane, the Camera cannot copy entire selections.

The data is copied to the clipboard or file in the Windows metafile format (WMF).

For more information on the Camera copy feature, refer to Chapter 3.

Alignment Pane

The Contig Alignment Pane (Figure 31.17) displays the nucleotide sequences of the fragments that form the contig with overlapping regions aligned appropriately and displayed relative to their positions in the contig. ORFs and translations can be identified and displayed. You can edit the sequences here and see how your actions are reflected in the contig alignment and consensus. The respective chromatograms for the sequences, also aligned appropriately, can also be dis-
played on command. Also, a Comparison Report can be generated between any 2 fragments, or a fragment and the consensus.

The Contig Alignment Pane has the following main elements:

- The **fragments names list** (at the left) contains all fragments that are used in this contig.
- The **fragment current position** is the relative position of the left end of the fragment (counted from the fragment start) that is currently shown.
- The **scale** (above the chromatogram alignments) gives a reference nucleotide position for all fragments and the consensus shown in the alignment.
- The **consensus sequence (nucleotide format)** shows the nucleotide consensus for the assembly.
- The **consensus sequence (symbol format)** shows the symbol consensus for the assembly. In the symbol consensus, blank spaces indicate identical residues, dots (•) indicate positions containing Ns and plus symbols (+) indicate ambiguity and/or gaps.
- When expanded (select fragment > Show Chromatogram), the **fragment track** (Figure 31.18) shows from top to bottom, an editable sequence, the non-editable sequence in dye-associated colors, and a chromatogram (if available and switched on). By hovering the cursor over a position in a fragment, a tooltip displays the original position of the base in the fragment. If the fragment has associated quality values, they are also displayed in the tooltip. This is useful in resolving an ambiguous position, if the outlier base has a low QV.
When either the Graphics or Alignment Panes is active, one of the fragments is always highlighted (by name in the fragment names list at the far left) and selected concurrently in other pane. To highlight another fragment, simply click on it with the left mouse button.

Figure 31.18 Details of a fragment track in the Chromatogram Pane

Finding a Sequence

A specific fragment sequence can be found in the Alignment Pane using the Find feature. Click on a fragment name in the Alignment pane list to select it. Launch the Find sequence dialog box either by choosing Edit > Find from the menu or by clicking the Find button ( ) on the toolbar. When the Find dialog box (Figure 31.19) appears, type in the sequence you want to find.

Click the Find Next button. If the sequence is found, the Alignment Pane displays the region where the sequence has been found and highlights the area. If gaps were introduced into the fragment being searched, they are ignored.

Identifying Ambiguous Positions in the Contig

The Alignment Pane can be configured for various text and background displays. Options are accessed on the Fragments Highlight tab from the View menu > View Options, or the respective Toolbar button ( ). The default settings (Highlight Subject = Ambiguity) and colors use a yellow background to highlight ambiguous positions, and red characters to identify the base calls in a fragment that differ from the contig consensus sequence at that position.

Find Next/Previous Ambiguous Symbol

To search a fragment or a contig consensus sequence for ambiguous positions or symbols ("N" or "R"), click on the Find Previous ( ) or Next Ambiguous ( ) buttons or select View > Find Previous or Next Ambiguous.

If you start this search with a sequence selected, the search proceeds from the selection in the specified direction. If there is no selection, the search starts from the start (Find Next) or the end (Find Previous) of the contig.
If an ambiguous position or symbol is found, its position is selected and displayed. If there are no more ambiguous symbols in the specified direction, you are informed as such.

Generating a Difference Report

You can generate a Difference Report between any 2 selected fragments, or a fragment and the consensus. Select on the fragment name in the fragment names list and use the Compare To... command in the View menu or shortcut menu. In the Difference Report dialog, select a target for comparison from the combo box menu and click Compare. The resulting Report (Figure 31.19) may be saved to a folder and opened in a text editing program.

![Difference Report](image)

Figure 31.20 Difference Report between a fragment and the contig consensus

Switch the Fragment’s Chromatogram On/Off

To alternate the display of the chromatogram of the highlighted fragment, click on the Show/Hide Chromatogram button( ) on the Pane Toolbar or select Show/Hide All Chromatograms on the shortcut menu. You can also select View > Show/Hide All Chromatograms from the menu bar. If the fragment does not have a chromatogram, these actions are disabled.

Repositioning a Fragment

Any fragment in the Alignment Pane can be moved vertically to bring it alongside any other fragment (or the consensus). Simply select the fragment name in the fragment names list, and drag it to the new position. (In contigs generated in Lite Contig mode, use the shortcut commands Move Fragment Up/Down.

Customizing Fragment Color in the Sequence Pane

You can customize the color of the text and its background in the Sequence Pane. For more information on customizing Sequence Pane text color, see Window Viewing Options, page 472.

Consensus Calculation and Display

The CAP3 algorithm calculates a consensus by weighted averages. The default display of the consensus sequence is according to the most probable base at a position, using QV scores when available. Alternative displays (i.e. using IUPAC ambiguity symbols) are available through
Positioning the Consensus

When the Contig Viewer is first opened, the consensus sequence (nucleotide and symbol formats) are displayed at the bottom of the alignment. This can be changed by clicking on the shortcut menu anywhere in the Alignment Pane to display both formats at the top of the pane, thus juxtaposing the consensus and the scale. Wherever the consensus is placed, any fragment can be positioned alongside it for comparison by dragging or by using the vertical scroll bars on the right side of the pane. Extended vertical scrolling from the Toolbar is available for contigs that contain a large number of fragments which cannot easily be viewed simultaneously on screen.

Translating the Consensus Sequence

The Alignment pane can display amino acid translations using standard translation, below (or above) the consensus sequence, in any or all forward reading frames, by toggling the respective buttons on the View menu or the Toolbar. Ambiguous positions in the consensus are translated as a single, defined amino acid if codon degeneracy permits (i.e. GGN = Gly). Gaps in the consensus are ignored in the translations, as they are in ORF calculation. The Toolbar button allows display in single-letter or three letter codes. To search the consensus for a reading frame that supports a specific conceptual translation, select the consensus, activate Set Translation Frame(s) from the View or shortcut menu, and type or paste in an amino acid string in single-letter code (Figure 31.21). You can also check/uncheck translation frames.

![Set Translation Frame(s)](image)

*Figure 31.21 Enter an amino acid string and search for a supporting reading frame*

Additional operations can be performed when the Contig Viewer is placed in Edit Mode. These are described in the next section.

Editing Contig Fragments or Consensus

Placing the Contig Viewer in Edit Mode from the Toolbar button allows you to edit the fragments that comprise the contig, change their positions or orientation or edit the consensus itself. In general, operations that alter the consensus sequence must be done in Edit mode. Obvious examples are insertions, deletion or replacements to a fragment or consensus, and shifting fragments or blocks. Some non-obvious operations are reverse complementing fragments or contigs, and declaring a reference sequence (the reference sequence is dropped from consensus calculation).

All editing changes in the fragments cause an immediate recalculation and redisplay of the contig’s consensus sequence. If necessary, editing changes in the contig consensus are reflected immediately in the fragment sequences that comprise the contig.
Although the basic sequence editing techniques are the same as in the Sequence Pane of the Fragment Viewer, there are some editing limitations and specifics because the fragments are incorporated in a contig. Since fragments in a contig can also be present in contigs from other assemblies in the project, such edits are allowed only after you confirm that you are aware that any other assemblies containing the fragment may be dismissed.

![Figure 31.22 Warning message regarding editing the contig](image)

**Note:** This warning only appears the first time you attempt editing changes (Figure 31.22). If you select **Yes**, you will be allowed to continue making more changes without warning. Click **Yes** to accept the changes. If you check the **Make contig read-only** box, the Yes option is disabled. If you select **No**, you are returned to the Contig Viewer.

If your changes are base deletions, the nucleotides are deleted from the chromatogram and the gap is closed. Bases to the right of the deleted region now move to the left and a vertical marker appears in the sequence at the position of the deletion. (If the base being deleted is at the left end of a fragment as positioned in the contig, there is an option of **not** moving the remaining bases to the left: Viewing Options > General tab > Keep position). The consensus is updated.

If the fragment track is displayed (with or without chromatogram), you can compare the edited sequence with the original sequence (dye-associated) below it. Also, if you open the edited fragment in the Fragment Viewer, the editing changes are reflected in the sequence.

To help you track edits in the contig, new symbols arising from insertions and replacements in the fragments are displayed in a teal color. Positions where a deletion occurred are indicted by a vertical red bar. These indicators are preserved for later reference when the edited contig is saved and closed.

**Note on Editing Lite Contigs**

When a lite contig is generated, the “fragments” seen in the contig are no longer distinct entities. Edits performed on sequences in the Contig Viewer are not transferred to the original fragments, but to local sequence copies that are part of the contig. Since the edited sequences are not distinct entities, they cannot be reassembled.

If you nonetheless want to reassemble the edited sequences, in the List Pane select all the fragments comprising the contig (not the contig itself). From the shortcut menu, choose Export Fragment > to GenBank or FASTA format. This will save the edited sequences in a multiple sequence file which can then be reloaded into Contig Express for further processing.

**Editing Commands**

The following operations require the Contig Viewer to be in Edit mode ( ). Once this is done, several Toolbar buttons become active. Many, though not all operations can be performed similarly in the Graphics Pane or the Contig Alignment Pane, with the results synchronized to the other pane. Note that positioning the cursor and making a selection are not identical actions and the following operations may require one but not the other.

**Insert Bases or Gaps**

In the Alignment Pane, position the cursor immediately ahead of the desired position in a fragment or the consensus and type one or more bases. IUPAC symbols are accepted. All symbols
are case-insensitive and always displayed in upper case. To insert a gap, use the keyboard space bar. A gap position is represented by a dash (-). Insertions in a fragment will update the consensus. Insertions made in the consensus will be updated to all fragments that map to that position in the consensus.

Delete Bases or Gaps

In the Alignment Pane, position the cursor immediately ahead of the desired position in a fragment or the consensus and press the Delete button on the keyboard one or more times. You can also select a continuous stretch of bases and use the keyboard Delete button to remove them. Bases and gaps are deleted in the same way. Deletions in a fragment will update the consensus. Deletions made in the consensus will be updated to all fragments that map to that position in the consensus.

Replace Bases or Gaps

In the Alignment Pane, select a single position in a fragment or the consensus and type the replacement from the keyboard. Replacement is disabled if more than one position is selected. Only IUPAC accepted symbols are accepted. Bases and gaps are replaced in the same way. However, replacing a base with a gap requires that the base first be deleted, and then a gap inserted in its place. A replacement in a fragment will update the consensus. A replacement in the consensus will be updated to all fragments that map to that position in the consensus. Note that at an ambiguous position in the contig, overtying the consensus with the current symbol is not a replacement and does not update disagreeing fragments.

Translating the Consensus

The consensus sequence translation can be displayed in the Alignment Pane in any one or all of the three direct frames. Display (or hide) the consensus translation in a particular frame by choosing the corresponding View > Consensus Translation > Frame <#> command from the Main Menu or by clicking its Consensus Translation - <#> Frame button counterpart.

Common Operations in the Graphics and Contig Alignment Panes

The following operations can be performed similarly in the the Graphics Pane and the Contig Alignment Pane.

Highlighting a Fragment

The following commands require that a fragment or block be highlighted. To highlight a fragment, click on a fragment “arrow” of a Contig in the Graphic Pane or on a fragment name in the Alignment Pane fragments name list. The same fragment is concurrently selected in other panes of the window.

Moving a Fragment

This action changes the position of a fragment in the contig. It is possible only when there are adjacent gaps in the direction of its movement or if the fragment relocation does not create any gaps in the contig.

There are multiple methods for moving a highlighted fragment in the contig

1. Press the keyboard CTRL button and continuously select a fragment. You will see a drag icon (_drag_ ) and you can drag the fragment to the new location. The new position of the fragment is shown in both a pop-up label and the workspace Status Bar. Release the mouse to drop the fragment into its new location. Fragment moving is cancelled if you switch to another window before dropping the fragment. Since moving a fragment is not allowed to introduce gaps into the contig, your ability to drag will be appropriately limited.
2. Select **View > Move Fragment** … or the corresponding command on the shortcut menu. In the dialog box that opens, enter the number of bases and direction (right or left). A text message calculates the permitted range of the move, given the rules described above. Click **OK** to move the fragment to the new position. The contig consensus is recalculated.

3. Click one of the **Move Fragment** buttons on the Toolbar. On two buttons ( ) one click moves the highlighted fragment one position in the respective direction. When movement in any direction is no longer possible, the respective button becomes disabled. The third button opens the dialog described above, to enter a range and direction.

For each of these actions, the consensus is recalculated and all panes of the Contig Viewer are updated. Use the **Undo/Redo** buttons to reverse or repeat each movement.

---

**Move a Highlighted Sequence Block in the Alignment Pane**

Blocks of sequence within a fragment may be moved left or right to close a gap. This is often done on blocks that are short enough to be viewed on-screen in their entirety. Highlight such a block in a fragment by double-clicking anywhere within it in the Alignment Pane. Click on the **Move block to the left/right** buttons ( ) to shift the block one position at a time. You can click repeatedly to move the block over several gapped positions. Once there is no longer a gap in that direction, the button is disabled. Moving blocks in the consensus sequence is not permitted.

---

**Reverse Complement a Fragment or Contig**

To reverse a fragment, highlight the fragment, click the **Reverse Complement Fragment** button ( ) from the Toolbar, or select **Reverse Complement Fragment** from the **View** or shortcut menu. The fragment will be reverse complemented, the contig consensus is recalculated and all panes of the Contig Viewer are updated.

To reverse complement the entire contig, use the command **Reverse Complement Contig** from the **View** or shortcut menu.

---

**Declare a Reference Sequence**

For use in specific situations, you can declare a fragment a reference sequence. The use of a reference sequence in a contig will be described more fully below. For now, it is sufficient to know that declaring a fragment a reference sequence is a form of editing, since the effect is to remove it from consensus calculation. The consensus is therefore recalculated for the remaining fragments in the contig. To make a reference sequence declaration, select the intended fragment and choose **Reference Sequence** from the context-sensitive shortcut menu.

---

**Use of a Reference Sequence**

For workflows in applications such as comparative genomics or forensics, it is useful to include a reference sequence in the assembly. Such a sequence may be introduced into ContigExpress in any supported file format and may or may not include a chromatogram. Add the reference sequence to the project in the Explorer List Pane and include it in the assembly. In the Contig Viewer, select the reference sequence in the Alignment Pane fragment names list, and click **Reference Sequence** from the shortcut menu. (This action is not permitted unless the Viewer is in Edit mode.) Only one sequence may be designated as reference at any one time in a contig.

The selected sequence reappears in lower case letters and pink highlight (Figure 31.23). Generally, it no longer contributes to the consensus calculation or any analysis derived from it (ambiguity, protein translation, etc.). However if the reference constitutes single coverage anywhere in the contig, it will define the consensus for that domain. The reference sequence can be
moved vertically in the names list to facilitate comparison with any other sequence, and in Difference Reports (page 482) it can be compared to any fragment or the recalculated consensus.

Figure 31.23 Reference Sequence AF043045 supports calling G at positions 2440 and 2442 of contig generated from fragments of HS3454_C02

Note: If any fragment in an assembly lacks QV scores (e.g. a .gb file intended as a reference sequence), QVs associated with all other fragments will be ignored during assembly. To avoid such omissions first exclude the reference from assembly, then re-assemble the resulting contig with the reference sequence.

Window Viewing Options

The following section summarizes various review and display options available for the Alignment Pane. Many of the default options have been described already in context. Other alternatives have been alluded to briefly or not at all.

The Contig View Options dialog box (Figure 31.24) for may be opened from the View menu > View Options, or by clicking on the Viewing Options button ( ) on the Toolbar. The Contig Viewing Options dialog box has five tabs for setting various Contig Viewer viewing parameters. You can save configured settings to a file using the Save button and load previously saved settings in using the Load button.

General Tab

By default, the Contig Viewing Options dialog box opens to the General tab (Figure 31.24):

- **Gaps** – Check the Do Not Count Gaps… check box if you want to ignore gaps in the assembled sequence when setting a selection or position via the Set Selection or Set
Cursor methods. This option is useful when you want to navigate the Viewer following the original numbering scheme of a fragment.

- **On Delete 5’ End** – Select the option to keep a fragment in its current aligned position or move it to the left when you delete bases from its left end as it appears in the contig.

**Fragment Highlights Tab**

The Fragment Highlights tab (Figure 31.25) controls custom highlighting and coloring options for the text and background in the Alignment Pane of the Contig Viewer.

![Figure 31.25 Contig Viewing Options dialog – Fragment Highlights tab](image)

- **Highlight Subject** – Determines the subject of highlighting/coloring that appears in the Alignment Pane.
  - **Bases** – Choose to color/highlight the text or background of individual bases in the Alignment Pane using chromatogram default colors (T – red, A – green, G – black, C – blue). Note that you cannot change the color schemes for the Text selection.
  - **Columns** – This provides a color scheme to columns in the alignment, based on several criteria:
    - **Ambiguity** is the default option. This provides a color background to any column with any base disagreements, and a counter color to the specific bases in that column that disagree with the consensus. A palette of colors is provided for the user to choose from, from the Basic Background Color and Text Color drop down menus.
    - **Probability** – Allows you to highlight the background based on the quality of the assembly at each position.
    - **None** – No coloring/highlighting of the sequence in the Alignment Pane.
- **Probability Highlight** – Controls color choice and dynamic range for the column highlight subject. The ATG graphic on the right provides a preview of how the current highlight configuration will appear in the Alignment Pane.
- **Brightness** – Controls the color brightness for minimum and maximum probability shading (i.e. choose to indicate higher probability with a higher or lower degree of shading).
- **Min/Max Probability (%)** – Designates the percent range that the ATG probability shading example pertains to. Narrowing the range displayed, allows for finer color gradation within the entire range.
**Weight Table Tab**

The Weight Table tab (Figure 31.26) allows you to change the nucleotide symbol weight coefficients used in calculating the Weight Graph data in the Graphics Pane.

![Weight Table tab](Figure 31.26 Contig Viewing Options dialog – Weight Table tab)

To change a coefficient for any nucleotide symbol pair, click in the table cell that is formed by the intersection of the nucleotide pair and enter a new value. The table has diagonal symmetry because the order of the symbols in the pair is irrelevant; the changes in one half of the table are instantly mirrored in the other half.

**ORF Tab**

In the ORF tab (Figure 31.27), you can define parameters for ORFs displayed in the Contig Viewer Graphics and Alignment Panes.

![ORF tab](Figure 31.27 Contig Viewing Options dialog – ORF tab)

Start and stop codons can be defined here, as well as the minimum codon size for the search, whether or not to include the stop codon in the ORF, and whether or not nested ORFs will be displayed. Clicking the **Defaults** button restores ORF parameters to their default values.
**Consensus Tab**

The Consensus tab (Figure 31.28) controls parameters affecting calculation of the contig consensus sequence displayed in the Alignment Pane.

![Consensus Tab](image)

- **Consensus Calling Algorithm** – Determines how ambiguous positions are displayed in the consensus.
  - **Most Probable Base** – The base call will be the most probable base at the position, determined by base frequency alone, or base frequency and quality values.
  - **IUPAC Nomenclature** – The base call will use the respective IUPAC codes designated for ambiguous positions (see Appendix C for a list of IUPAC conventions).
  - **Stringency** – The base call will be N if the **Set Probability Threshold** (see below) is un-checked. If checked, it will call the most probable base when the probability is above the set threshold, but call N if it is below.

- **Determine Probability By** – Sets the method used for determining probability. Probability can be based on frequency of occurrence or on base frequency and Phred quality values (see Importing Phred Quality Values, page 416). If the latter is checked but there are no such values in the fragments, determination will be by most frequent occurrence.

- **Set Probability Threshold** – For Stringency and IUPAC Nomenclature calling algorithms, allows you to set a threshold, above which the most probable base is called, but below which N is used.

**Copy and Camera Features**

The Copy feature is the same as in other panes except that only the selected part of a highlighted fragment is copied.
The Camera feature for this pane works in a slightly different manner than it does in other windows. When you select this action, you will see the Camera dialog box (Figure 31.29):

![Camera dialog box](image)

Figure 31.29 Camera dialog

In the Metafile format, only the currently visible part of the Alignment Pane can be copied to the clipboard. The data is copied in the Windows metafile format (WMF).

In the Text format, you can select whether you want to copy a current selection or all data available in the contig (it is not limited to the visible part). You can also set wrapping of the copied sequences at every nth position. The data is copied in the plain text and Rich Text (RTF) formats.

Check the **Exclude Consensus** text box to exclude the contig sequence from the copied text in the Sequence Pane (this feature does not exist in the Graphics Pane Camera dialog box).
Introduction

This chapter consists of operations performed on fragments to optimize contig assembly. Operations covered are fragment and vector contamination trimming and calling secondary peaks.

Fragment Trimming

Fragment trimming is an operation performed on chromatogram fragments to optimize sequencing results and contig assembly. Trimming is performed on fragment ends to remove unresolved or poor quality nucleotides based on chromatogram results. It is also used to remove bases identical to restriction sites or other sequences from a vector. An entire chapter is devoted to this operation because of its complexity.

1. All fragment trimming is a multi-step process in Vector NTI Advance. Steps in the process are summarized as follows:

2. Highlight the fragment(s) you want to trim.

3. Open the Fragment Ends, Vector Contamination or Phred Quality Values Trimmer dialog box where trimming parameters are specified.

4. Define the trimmer settings.

5. Calculate the trimming using the parameters you defined. For Fragment Ends and Vector Contamination trimming, a preview is provided so you can see where the fragments will be trimmed.

6. Select or unselect all of the ends to be trimmed.

7. Complete the trimming process by closing the dialog box with the OK button.

Fragment Ends and Vector Contamination Trimmer Dialog Boxes

In the ContigExpress Project Explorer, select the fragments to be examined and prepared for trimming. To trim selected fragment ends, choose Edit > Trim Selected Fragment Ends from...
the Main Menu, opening the Fragment Ends Trimmer dialog box (Figure 32.1). To trim selected fragments ends for vector contamination, choose Edit > Trim Selected Fragments For Vector Contamination from the Main Menu (Note: the Vector Contamination dialog box is identical to the Ends Trimmer dialog box except for the title).

When first opened, the dialog box displays the default settings for the type of trim used. While the dialog box is open, you can change the trimming settings and recalculate the trims as many times as you wish.

The dialog box contains a scrollable list of Fragment mini-windows for all the fragments you selected for trimming in Project Explorer. To facilitate viewing the trim locations, this dialog can be resized like any other window using the keyboard or mouse.

![Figure 32.1 Fragment Ends Trimmer dialog box](image)

The Fragment Trimmer dialog box buttons are described in the steps below. It is suggested that you follow them as numbered.

The fragments you selected in the Project Explorer are listed in “mini-windows” in this dialog box. Each mini-window shows one fragment with information on how the current trim will affect it. In this window, you can select those trim locations to be used in the final trim for the fragment in the project.

When the fragments are displayed as sequences (the Display Sequence box is checked), the Fragment mini-window (Figure 32.2) has the following elements. (Some of the elements do not appear until after the calculation step):

![Figure 32.2 Fragment Mini-window when fragments are displayed as sequences](image)
- **Fragment name** is the name of the fragment in the ContigExpress Project;
- **Fragment statistics** show the number of good (non-ambiguous) bases and the number of ambiguous bases in the fragment;
- **Fragment sequence** is the nucleotide sequence of the fragment; the ends that will be trimmed are in lowercase and selected in the red color, all the rest will remain intact after trimming;
- **Trim statistics** shows the amount of bases that will be trimmed off the fragment from each end and the amount of ambiguous bases among them;
- **Select trimming location boxes** allow you to select the ends that will be trimmed.

Due to window size constraints, usually only the most important parts of the sequence, the places around the trim locations, are displayed. Omitted parts of the sequence are replaced with ellipses (...). To view the entire fragment in a Fragment Display Window (in Read-Only mode), double click anywhere in a mini-Window.

To display fragments in a graphical form, with a bar differentiating (in different colors) the parts for trimming from the remaining sequence part as seen below, uncheck the Display Sequence box. The result is shown below (Figure 32.3):

![Fragment Mini-window when fragments are displayed in graphical form](image)

**Figure 32.3 Fragment Mini-window when fragments are displayed in graphical form**

Follow the trimming steps in the numbered order:

1. **Set fragment display options**
   
   To establish the fragment display format, check or uncheck the Display Sequence box or press alt-D.

2. **Define trimmer settings**
   
   Click on the **Settings** button or press alt-S, opening the Trimming Settings dialog box. The contents of this dialog box may vary according to the type of the Fragment Trimmer dialog box. See the section on Vector Contamination Trimming for more details.
**Fragment Ends Trimming Settings dialog box**

The settings defined in this Trimmer Settings dialog box (Figure 32.4) determine how contig fragments’ ends are trimmed. When the dialog box is opened, its fields are filled with the current settings. To load previously saved parameters, click the Load from File button.

![Fragment Ends Trimming Settings dialog box](image)

5’ end:
- **Trim __ consecutive off-scale bases** removes the defined # of consecutive bases that are below acceptable criteria
- **Trimming until __ 5’ bases** can be based upon chromatogram quality
- **Trim at least __ 5’ bases** is an arbitrary setting that may be based upon the fact that your primers have “tails”

3’ end:
- **Trim until __ 3’ bases**... % of maximum value removes bases whose peaks do not meet the % value you define
- **From position ... trim first**... removes these bases if you can see that quality well into the sequence is not good
- **Trim until __ 3’ bases**... ambiguities trims poorly resolved bases
- **Trim at least**... removes bases that begin to be poorly resolved at the 3’ end

Post trimming:
- **Define maximum**... set the length of the fragment that must be left after trimming
- **Leading and trailing**... removes poorly resolves nucleotides that may be left after trimming
- **Remove poly A/T**... removes these nucleotides that may be present if the sequence was flipped, producing a poly-T 5’ end.
Click **Save to File** to save these settings as a file to be used in the future. Click **OK** to apply the new settings in the current trimming exercise and return you to the Fragment Ends Trimmer dialog box. This also saves these settings as default settings for all future trim calculations. To cancel changes, click the **Cancel** button.

If you calculated the trims before redefining these settings, you must recalculate after closing the Settings dialog box to view trimmings according to the new settings.

3. Calculate the trim

**Calculate!** shows the potential trimming under the parameters you set in the Trimmings Settings dialog box. To calculate the trim for all fragments using the current settings, click the **Calculate!** button or press alt-C. Since trim calculation can be a lengthy process, a progress dialog box appears with a progress bar (Figure 32.5). **Click the Cancel button or the ESC key to stop the process.**

![Figure 32.5 Progress dialog box with progress bar](image-url)

If the trim calculation process is allowed to complete, all fragments are updated with newly calculated trim locations, displayed in red, lower-case nucleotides in the mini-windows (Figure 32.6).

![Figure 32.6 Calculated trims to remove ambiguous regions from fragment ends](image-url)

4. Select trimming locations

Click **Select All Ends** or press Alt-A to select all of the trim locations displayed in the mini-windows for the final fragment trimming. Click the **Select All Ends** or **Unselect All Ends** buttons or use keystrokes Alt-A or Alt-U respectively. Note that all of the checkboxes for the ends are checked as in this example:

![Selecting trim locations](image-url)

When a location is selected, the respective part of the fragment will be trimmed off when the trimming is completed.
5. Confirm the proposed trimming operation
   To open a Fragment Window for any fragment in a mini-window so that you can review and
   compare prospective fragment trimming with the corresponding chromatogram, double-click
   anywhere in the mini-window. The Fragment Window that opens is in read-only mode.
   As this window is created, the fragment that will remain after trimming is selected in the frag-
   ment sequence, both in the Sequence Pane and in the top sequence of the Chromatogram
   Pane. The fragment opened this way is always opened in read-only mode.
   If the Fragment Window for the fragment already exists, the selection is made inside it and it is
   brought to the foreground.

6. Complete the trimming operation
   Once parameters and trimming conditions are established, click OK to close the dialog box and
   proceed with the fragment trimming. To close the dialog box without changing the fragments in
   the project, select the Cancel button or press the ESC key.

Vector Contamination Trimming Settings Dialog Box
   The purpose of the Vector Contamination Trimming feature is to trim from sequenced fragments
   any residues from your frequently used cloning vectors that may have been amplified in the
   sequencing process.
   In the Vector NTI Database Explorer, highlight the cloning vector in the List Pane, or open a vec-
   tor in a Molecule Display window and select a sequence such as a frequently-used polylinker
   site. Select Tools > Send to > Polylinker to ContigExpress.
   In the Choose Polylinker file name dialog box, name the sequence, making sure you retain the
   .seq extension because this is the only way ContigExpress recognizes this sequence for the
   trimming operation. Click Save to save the sequence in the VecContData folder.

Notes about sending sequences for vector trimming:
   • From Database Explorer, you can send only one vector at a time.
   • If you have been sequencing off of the complementary strand, be sure and indicate this,
     as only one strand is forwarded to ContigExpress.
   • In naming the sequence, make sure you retain the.seq extension because this is the
     only way ContigExpress recognizes the sequence for the trimming operation.
   Now in the Project Explorer in ContigExpress, select the sequenced fragments to be trimmed
   and choose Edit > Trim Selected Fragments for Vector Contamination. This opens the Frag-
   ment Vector Contamination Trimmer dialog box, (not shown but similar to the Ends Trimmer dia-
   log box).

Vector Trimmings Settings
   In this dialog box, click the Settings button to open the Fragment Vector Contamination Trim-
   mer Settings dialog box (Figure 32.7). The vectors or polylinker sites you selected in the Data-
   base Explorer should be listed in the List Pane at the left. The Edit Pane where you can edit the
   vector sequence is on the right.

Tip: If when you open this dialog box, the Polylinker List Pane is empty, there can be two reasons/
   solutions:
   1. You may have defined it incorrectly. Return to the Project Explorer, and select your vector(s)
      or polylinker(s) as described above.
2. You may have selected the vector/polylinker correctly, but you lost the .seq extension in naming them when they were saved in the VecContData folder.

The Common Settings in this dialog box are described as follows:

- **Minimum vector overlap** is the minimum number of bases in the fragment that overlap with those on the clone. *This setting must be 5 or greater.*
- **Minimum vector overlap with ambiguities** includes poorly resolved residues
- **Vector match threshold** is the % required to match
- **Remove additional ... from 5’/3’ end...** refers to the additional bases to be removed

The polylinker list in the Polylinker List Pane on the left is updated every time you open this dialog box. Check the Auto-refresh box to refresh the list every time you switch to ContigExpress from another application where the Vector Contamination Settings dialog box is open.

By default, a new polylinker is not selected for trimming and has no insertion points.

**Editing a Polylinker**

To select a polylinker for the trimming process, check the check box to the left of its name and highlight it by clicking on it. To Remove a Polylinker from the list, click the Remove Polylinker(s) button. The Polylinker Edit Pane to the right of the list shows the contents of the selected polylinker. You can use this window for simple editing of the polylinker sequence and for managing the insertion points of the polylinker. Editing commands appear on the buttons; some are also available on a shortcut menu opened from the Edit Pane.

Insertion points can be added to the polylinker using two methods:

- **Manual adding:** position the caret where you want to add the insertion point; press the INSERT key or click the Add Insertion Point button.
- Import a group of insertion points using a file with an enzyme database (in Rebase format). Click the **Add REN sites** button or select **Add REN Sites...** from the shortcut menu. In the standard Open dialog box, you can select a Rebase file to be applied to the polylinker or simply the Enzlist.dat in the VecContData folder. Select **Open**, and the enzyme file is imported. The insertion points with the enzyme names are imported to the polylinker. Click on the enzyme name to select it as an insertion point.

Selecting on an enzyme label executes a trim up until (or from) the enzyme cut site. If you wish to trim the site altogether, use the site display as a guide and manually place the insertion point after (for 5' contamination) or before (for 3' contamination) the recognition string.

Vector NTI Advance 10 comes with several .dat files: a preselected set of common enzymes for restriction cloning, a set for TOPO cloning, and for Gateway cloning. In addition, a blank file (MySites.dat) is provided for you to introduce your own cloning sites. To use, open this file (C:\VNTI Database\VecContData folder) in NotePad, follow the comments and Save.

The following options control the points:

- **Remove Insertion Point** deletes a selected point. You can also press the **DELETE** key.
- **Remove All Points** (on the shortcut menu) remove all insertion points from a polylinker
- **Select All** selects all insertion points. You can also just click on it. (Selected points are blue; an unselected point is gray.) A corresponding option is also on the shortcut menu. To be included in the trimming process, a point must be selected.
- **Remove Selected** deletes selected points. A corresponding option is also on the shortcut menu.

To toggle the selection status of an insertion point, click it with the mouse.

Click on the **OK** button to return to the Trimmings dialog box.

**Executing the Vector Contamination Trim Changes**

To complete the trimming process, click Calculate. When the trim is calculated, the polylinker sequence and Insertion Points and their reverse complements are automatically used, so that forward and reverse sequencing reads can be trimmed using one setup.

After all fragments are processed, they are available in the Fragment Trimmer dialog for review (Figure 32.8). When satisfied with the results, dialog can be closed by clicking **OK** and the Project Explorer contents are updated to reflect the changes.

**Figure 32.8 Trim of vector contamination from reads of HS354_CO2. Fragments were cloned in Gateway vector pENTR221**

Additional notes on vector trimming:

1. Trimming contamination from TOPO cloning projects.
a. TOPO vectors are represented in the Vector database as they are supplied and used, i.e., as linear molecules. In this form, it is impossible to define a poly linker crossing both cloning termini. We recommend temporarily converting the vector to circular (DNA/RNA menu > Edit > DNA/RNA Molecule tab) in order to define a suitable poly linker.

b. Many TOPO vectors are highly symmetric about the 2 cloning termini. Use of the default trimming parameters may trim the entire read. We recommend values of $\geq 15$ for the overlap size and $\geq 85$ for the match threshold.

2. Trimming intronic sequence from sequencing reads of amplified exons.

You can use vector trimming operations to remove intronic sequence from amplified exons. In these workflows, the sequencing reads are often primed from sites in the intron regions. To do this, it is necessary to create a poly linker that joins together direct strand sequence from before and after the exon under investigation.

a. Using VNTI, open a file containing the annotated exon bounded by intron sequence.

b. Make a continuous selection from 50 bases before to 50 bases after the exon, i.e. the exon as well as 5’ and 3’ intron sequence.

c. Use the selection to create a new linear molecule from the Goal List (Add Fragment to Goal List > Construction fragment > Finish > Run > Construct). Name it exon poly linker.

d. Open the exon poly linker file, select the exon and delete it; the exon poly linker is now 2 concatenated intron sequences. Save.

e. Send the edited exon poly linker file to Contig Express as a poly linker (Tools menu > Send To > Poly linker to Contig Express).

f. In the Vector Trimming dialog, edit the poly linker by manually placing an Insertion Point exactly where the 2 intron sequences are joined. If the poly linker file was initially defined from -50 to +50 (step 2 above), you can simply place the Insertion Point at the very end of the first sequence row.

Calling Secondary Peaks

ContigExpress can search for secondary peaks in a chromatogram by looking for lower peaks that are as tall as a percentage you define of the highest peak at a base call position (current threshold for secondary base calling). For example, a threshold of 70% would display secondary bases with heights 70% of the taller peak as the appropriate IUPAC ambiguity symbol. This is specifically used for resolving the occurrence of double peaks in a chromatogram.

To initiate this process, in the ContigExpress Project Explorer window select fragments, then select Edit > Call Secondary Peaks for Selected Fragments. This opens the Call Secondary Peaks dialog box (Figure 32.9) where you can review potential replacement bases determined by the search for secondary peaks for the group of fragments you selected.

To facilitate viewing the fragment sequences, this dialog can be resized like any other window by using the keyboard or mouse. The panes inside the dialog may be resized by moving the split bars.

The dialog box opens with the default settings for this operation which you can change as you wish. The dialog box has two panes: List Pane on the left with the fragments you selected in the Project Explorer and the Sequence Pane on the right displaying the sequence of the currently selected fragment in that list.

List Pane

The List Pane shows the list of fragments you selected in Project Explorer for sequence editing. Select any fragment in the list to see a preview of the changes in the Sequence Pane. If you are
satisfied with the results, check the box next to the fragment name to allow the displayed changes for this fragment to be saved in the project.

**Sequence Pane**

This pane displays the preview of changes that will be applied to the sequence using the current call peaks settings. The peaks that will be changed are shown in blue color. The original peaks are shown right under the changed peaks so you can easily track changes (Figure 32.9).

![Figure 32.9 Call Secondary Peaks dialog box](image)

The Call Secondary Peaks dialog box buttons are summarized as follows:

- **Allow Ns to be replaced** allows replacing the Ns with less ambiguous IUPAC symbols, if possible.
- **Allow edited bases to be replaced** allows replacing of previously edited bases.

If either of the above settings is checked, the selected sequence is immediately regenerated to display a preview of the base that would be changed in the sequence.

- **Threshold for secondary base calling** defines the percentage of the height of the secondary peak in relation to the first for calling the secondary base
- **Select all fragments** all fragments with secondary base replacements will be saved into project after clicking OK.
- **Unselect all fragments** reject changes to all listed fragments resulting from recommended secondary calls.

**Status Bar**

The status bar of the Call Secondary Peaks dialog box (Figure 32.10) shows the name of currently selected fragment in List Pane, and the number of peaks that have changed in its sequence.

![Figure 32.10 Call Secondary Peaks status bar](image)
The Set Caret Position and the Set Selection Range boxes show the caret position and a currently selected range of residues in the Sequence Pane. Pressing either of these buttons opens a dialog box where you can set the caret position or range directly.

After you have previewed the changes, check the check box to maintain the changes while you review other fragments. To select other fragments to be changed, click on their fragment IDs in the List Pane.

**Executing Calling Peaks Changes**

The changes are applied for those fragments checked in List Pane only. To save your changes in the project, click the **OK** button. If any of the affected fragments are included in assemblies, you will be given the option of dismissing the respective assemblies or canceling the changes for that particular fragment.
Introduction

GenomBench, Vector NTI Advance’s genome project application, can be used for downloading, viewing, analyzing, annotating and saving local copies of reference genomic DNA sequences served by several principal Distributed Annotation System (DAS) servers as well as private DAS servers. Like Vector NTI, GenomBench has a local database for storing your sequence and analysis results.

GenomBench was designed to support the following workflows:

- Retrieval of genomic data from public DAS servers.
- Graphical and workflow support for editing transcripts under genomic context to obtain full-length cDNAs.
- Positioning or mapping custom transcript sequences onto a genomic backbone.
- Annotating and creating a local copy of a genomic region.

In support of these workflows, it accepts megabase-sized genomic sequences, and all attendant track annotations, and displays the information in an integrated, graphically rich window that provides meaningful access to user-selected regions of human and other genomes. All annotated features are linked to their corresponding region of the genomic sequence backbone; additional information for any sequence-based annotation (for example, mRNAs, ESTs, and STSs), including other defined features from the GenBank record, is easily downloadable. Annotated sequences can be aligned with the genomic backbone simply by dragging-and-dropping into an alignment window.

Proprietary DNA sequences, stored in the Vector NTI local database, can also be positioned along genomic backbones using Spidey or Sim4 alignment algorithms to determine the intron-exon structure of genes of interest. Moreover, genomic sequences obtained, analyzed and annotated in GenomBench can be exported to Vector NTI for further sequence analysis, such as PCR primer design or the creation of spliced transcripts.

GenomBench also provides the capability to query public databases, perform sequence and feature searches, translations and alignments, and copy and paste sequence and graphics.
Launching GenomBench

To launch GenomBench, select Start > Programs > Invitrogen > Vector NTI Suite 10 > GenomBench, assuming a default Vector NTI Advance installation has been performed. You can also double-click its icon, NauViewer.exe, in the program group or folder in which you installed Vector NTI Advance. When GenomBench is first opened, the GenomBench window is empty. GenomBench can also be launched from the Analysis Monitor workspace for completed Analysis Results. See Viewing Analysis Results in GenomBench on page 572.

GenomBench provides local storage capability in the form of a database file or .mdb file. If you did not previously have GenomBench installed, continue to the next section to configure database paths. If you had the previous version of GenomBench (version 1) installed on your system with Vector NTI Advance (version 8), the first time you open GenomBench (version 2) with Vector NTI Advance (version 9) the message in Figure 33.1 displays asking you to migrate your current database so that it is compatible with the new version of GenomBench.

If you plan to use an existing version of a GenomBench database with the previous version of GenomBench and also want to use that database with the newer version of GenomBench, click Cancel in the message, and make a backup copy of the database(s). To locate all copies of existing databases, select Start > Programs > Invitrogen > Vector NTI Suite 10 > GenomBench Utilities > LS Configuration. Paths to all configured databases are listed in the drop-down menu in the Database Name field. To backup the database, simply copy the .mdb file to a new location.

To continue with the migration, click OK in the message. The migration proceeds and opens the GenomBench window when finished. Any existing data remains in the database and you can now save new data as well.

Configuring Database Paths

GenomBench can connect to its own local database and to DAS 1.0/1.5 compliant servers. Your Vector NTI Advance (version 9) installation includes two utilities, LS Configuration and DAS Configuration, for configuring paths and connecting to local GenomBench databases and to public DAS compliant servers, respectively. Although Vector NTI and GenomBench can exchange data, the databases are independent for each.

**Configuring the Local Database**

The first time you open GenomBench, if there are no previously created local GenomBench databases, you must launch the LS Configuration utility to create a local database. To launch this utility, select Start > Programs > Invitrogen > Vector NTI Suite 10 > GenomBench Utilities > LS Configuration. The Local Storage Configuration dialog box (Figure 33.2) opens...
where you can create local GenomBench databases or browse to connect to previously created local databases.

Figure 33.2 Local Storage Configuration dialog box

- **Database Name** – Enter the name of the new local GenomBench database you want to create or click the **Browse** button to locate an existing local GenomBench database you want to connect to. A suggested name for the default database is provided called `gb.mdb` located in `C:\VNTI Database`.

- **Database Type** – Currently only the Microsoft JET database type is supported.

- **User** – User name for the database (not required for JET).

- **Host** – Host name for the database (not required for JET).

- **Port** – Port designation for the database (not required for JET).

- **Password** – Password for the existing database (not required for JET).

- **Reset** – Resets the configuration dialog box to the default folder, `C:\VNTI Database\gb.mdb`.

- **Create DB** – Creates the newly configured database.

- **Connect** – Connects to the specified, existing database.

For new GenomBench installations, the Local Storage Configuration dialog box shows a suggested path/name in the **Database Name** field the first time it is launched. Clicking the **OK** button creates this default local database.

The **Connect** button is generally only used when you are connecting to a newly created database or when you are changing the configuration to view a different local GenomBench database that you have previously created.

- You can create multiple local databases and use the LS Configuration utility to connect to them/direct GenomBench to view them.

- Once you have connected to a local database with the LS Configuration utility, GenomBench is directed to that database.

- GenomBench can view only one local database at a time.

**Configuring DAS Servers**

Five public database DAS servers are already configured for you in GenomBench. These are UCSC, Ensembl, euGene, WormBase, and TIGR Arabidopsis. You can view the list of configured DAS servers and their configuration parameters or configure the parameters for a new DAS server in the DAS Configuration dialog box (Figure 33.3). To open the DAS Configuration
dialog box, select **Start > Programs > Invitrogen > Vector NTI Suite 10 > GenomBench Utilities > DAS Configuration.**

Figure 33.3 DAS Configuration dialog box

Highlight a server name to view or edit its configuration parameters.

To configure a new server, click the **New server** button. An entry is added for the new server. Configure the parameters for the new server as follows:

- **Server Name** – Click the Server Name field for the new entry and enter a server name.
- **URL** – Click the URL field for the new entry and enter the URL.
- **Fragment Size Limit** – Identifies the maximum size fragment this server allows for download at one time; larger fragments are downloaded in appropriately sized chunks. For example, the UCSC server has a fragment size limit of 200K. To set a fragment size limit, select the **Fragment Size Limit** box and enter the limit size in the textbox.
- **Type Name for CytoBand Features** – Identifies the feature type name for this server for cytoband features. Of the pre-configured public DAS servers, UCSC, EuGene, Wormbase, and TIGR have cytoband features; Ensembl does not.

To delete a server, highlight its name and click the **Delete server** button.

Click the **Save** button to save any changes.
Opening a Sequence from a DAS Server or a Local Database

Genomic sequences can be opened either from the configured DAS server databases or from a local GenomBench database.

To open a genomic sequence in GenomBench, select File > Open or click the Open button ( ) on the toolbar. In the Open Molecule dialog box (Figure 33.4), select the DAS or Local Storage tab to open a genomic sequence or saved fragment, respectively.

![Figure 33.4 Opening a genomic sequence from a DAS server](image)

In the DAS tab (Figure 33.4), select the server from the drop-down menu and highlight the appropriate source in the Subset column. Select the molecule from the Name column and click OK.

In the Local Storage tab (Figure 33.5), highlight the appropriate subset in the Subset column and select the fragment from the Name column. Click OK.

![Figure 33.5 Opening a saved fragment from the Local database](image)
**Note:** The local GenomBench database is empty until you save sequences to it. For more information, see Saving Sequences on page 552.

Once a genomic sequence is opened in GenomBench, information related to the sequence displays in the Overview Pane and the Text tab of the Information Pane (Figure 33.6), but the Feature Map Pane and Sequence, Alignment and Features tabs of the Information Pane remain empty until a fragment is loaded. If the DAS server supports cytoband features, these display in the Overview Pane. The title bar shows information about the downloaded sequence.

![Figure 33.6 Genomic Sequence opened in GenomBench; no features added](image)

**Loading Features into the Feature Map Pane**

To load features from a DAS server for a fragment into the Feature Map Pane, highlight the fragment area in the Overview Pane and do one of the following:

- Select **Load Highlighted Range** from the shortcut menu.
- Click the **Load Fragment** button ( ).
- Select **File > Load Fragment** from the drop-down menu.

The features of the highlighted fragment are loaded. If you select **Load Fragment** from the File menu or toolbar button, the Load Fragment dialog box displays, allowing you to define the fragment by start position and length. Click **OK** to load the features.

**Note:** You can set a maximum size of fragment to download. If your selection exceeds the maximum size, a warning message displays asking if you want to continue and load the first part of the fragment up to the maximum size set or if you want to load in Sequence on Demand mode, which downloads sequence as needed. For more information, see Editing Fragment Loading Configurations on page 534.

If features for the designated fragment exist on the server, the feature information loads into the Feature Map Pane and the Sequence, Text and Features tabs of the Overview Pane in the GenomBench window (Figure 33.9). Features load from the bottom to the top of the Feature Map Pane, progressively, and are categorized by type. Feature type names display on the left side of the Feature Map Pane.
While features are loading, the left icon in the lower right corner of the status bar flashes to show the progress of loading. Clicking on the icon opens the Data Loading Monitor (Figure 33.7).

![Data Loading Monitor](image)

*Figure 33.7 Data Loading Monitor shows the loading status of features*

The **Main Sequence** tab in the Data Loading Monitor shows the loading status of features in the fragment. If the Status column contains 'in queue', the feature in the Request column has been downloaded from the server and is waiting to be loaded. The Time column displays the time the features were downloaded from the server. As features are successfully loaded, they are removed from the list. If you do not want to load a particular feature, select it and click the **Abort** button. Features that cannot be loaded remain in the Main Sequence tab and display a status of failed. If there are features that failed to load, the icon in the lower right corner of the status bar changes from ![progress](image) to ![failed](image) to indicate that one or more features could not be loaded. For information on why the feature failed to load, select the feature in the Data Loading Monitor and click the **Details** button.

The **Supporting Information** tab in the Data Loading Monitor shows the loading status of target sequences when features are added to the Alignment Tab (Figure 33.8). If a feature has alignment information, then GenomBench can download the feature’s associated sequence (usually) from NCBI so you can use it for alignments. For more information, see **Alignment Tab** on page 523. As target sequences for selected features are loaded into the Alignment Tab, the status of the loading is shown. The right icon in the lower right corner of the status bar also shows the progress of target sequence loading. Loading status also displays when features are transferred and when a target sequence is opened. For more information, see **Target Sequences** on
Similar to the Main Sequence tab, the Supporting Information tab has Request, Status, and Time columns and Abort and Details buttons.

![Figure 33.8 Supporting Information tab of the Data Loading Monitor](image)

You can designate the types of feature to load by Setting Feature Type Filters, described on page 530, and Editing Fragment Loading Configurations, described on page 534. It can take some time for all the features to load, but you can start working on features that have loaded before the remaining features are loaded. When the feature loading icon stops flashing ( ), feature loading is complete.

![Figure 33.9 Genomic sequence opened in GenomBench with features loaded for selected fragment](image)

A mark displays in the Overview Pane ( ), indicating the position on the genomic sequence of the fragment currently shown in the Feature Map Pane. The Sequence, Features and Text Tabs
in the Information Pane display data for the selected fragment. In addition to genomic sequence information, the GenomBench title bar now shows fragment information.

**Features Downloaded from Ensembl**

Because of the way the DAS protocol is implemented at the Ensembl site, when fragment features are loaded from Ensembl they come in as one condensed feature called a *component*. To extract features from *components*, perform the following steps.

1. Click the + symbol on the left side of *components* in the Feature Map Pane or Name column in the Features tab of the Information Pane to expand the subfeatures. Likewise, if + symbols exist to the left of any of the extracted features, clicking this symbol further extracts these features.

2. In the Feature Map Pane, highlight the features you want to extract. You can highlight multiple features by pressing the CTRL key on the keyboard while clicking and dragging a box around the features.

In the Feature Map Pane, drag and drop the selected features above the main *components* feature. Subfeatures can now be grouped according to type (see Feature Grouping/Ungrouping Operations on page 520).

**Features of the GenomBench Window**

**General Pane Operations**

A number of basic operations can be performed in the different GenomBench panes, including activating panes, removing/restoring panes from view, setting selections and fragments, marking/unmarking selections, and various zooming operations.

**Activating Panes**

Panes are activated in GenomBench by clicking on them, as they are in other Vector NTI Advance applications. You can also activate a pane (or tab in the case of the Information Pane) by selecting the appropriate command from the View menu: View > <pane or tab name>.

Panes in GenomBench are labeled with the Pane name on a vertical bar on the left border of the pane. The active pane label is blue/highlighted; pane labels for the inactive panes are grey. Only one pane can be active at a time.

**Removing/Restoring Panes**

The Overview and Information Panes can be removed from the GenomBench window. The Feature Map Pane cannot be removed from view.

To remove a pane from view, close it by clicking the X at the top of the pane’s title bar.

To restore a removed pane to the GenomBench window, select View > <pane name> from the drop-down menu.

The Overview and Information Panes can also be removed from the GenomBench window by dragging them by their title bars and dropping them outside the GenomBench window or by double-clicking their pane’s title bars. In this case the pane becomes “undocked” from the main GenomBench window and can be moved independently as a separate window. You can re-dock the pane to the GenomBench window by dragging the pane back into the GenomBench window, above or below the Feature Map Pane, or by double-clicking the pane’s title bar again.
**Zooming Panes**

Zoom the Overview and Feature Map Panes by activating the pane and choosing one of the zoom options from the View menu or by pressing the corresponding zoom button on the toolbar. Panes are zoomed independently. Zoom functions are described in detail in Chapter 3.

**Highlighting Sequence Regions**

Regions can be highlighted in the Overview and Feature Map Panes and in the Information tab of the Sequence Pane by clicking and dragging with the mouse, similar to other Vector NTI Advance applications.

In the Overview Pane, a whole sequence, a range of the sequence, or the current fragment can be highlighted by selecting the corresponding commands from the Overview Pane shortcut menu.

**Setting Selections and Fragments**

GenomBench also provides a Set Fragment feature for defining a selection according to fragment coordinates in the Overview Pane and a Set Selection feature for defining a selection based on coordinates in the Feature Map Pane and Sequence tab of the Information Pane.

To set a fragment or selection, select **Edit > Set Fragment** or **Set Selection**, respectively. In the Set Fragment or Set Selection dialog box, enter the Start position and Length for the selection and click **OK**. The corresponding selection appears in the Overview or Feature Map Pane and Sequence Tab.

You can also set a selection by double-clicking a feature. The region spanning the feature is selected.

**Marking and Unmarking Selections**

Regions of interest and features can be marked and unmarked in the Feature Map Pane and in the Sequence tab of the Information Pane. Marking places a vertical, shaded area along the genomic backbone and helps to delineate all the features in the region (Figure 33.10). Marking a feature can be a powerful method for distinguishing portions of interest in the feature. For example, marking a selected gene prediction feature marks all the exons of the feature but not the introns.

![Figure 33.10 Selected acembly feature showing marked exons and unmarked introns](image)

To mark a region, select it in either the Feature Map Pane or Sequence tab of the Information Pane using one of the selection techniques described above and click the **Mark Selected Segment** button or select **Marks > Mark Segment** from the Edit or shortcut menus.

To unmark a region, select it and click the **Unmark Selected Segment** button or select **Marks > Unmark Segment** from the Edit or shortcut menus.

To remove all marks, and click the **Remove All Marks** button or select **Marks > Remove All Marks** from the Edit or shortcut menus.
To mark a feature, click on it in either the Feature Map Pane or Sequence tab of the Information Pane to select it and select **Marks > Mark Selected Feature** from the Edit or shortcut menus. To unmark a marked feature, select **Marks > Unmark Selected Feature** from the Edit or shortcut menus.

When features or regions are marked, you can jump from one marked region or feature to the next by selecting **Edit > Select Marked Segments > Next** or **Previous** or use the shortcut keys > and < to move to the next or previous mark respectively. The current marked region is highlighted in the Feature Map Pane as well as in the Sequence Tab of the Information Pane.

**Overview Pane**

The Overview Pane presents a graphical depiction of the genomic sequence (Figure 33.11) and has the following features:

- A rod-shaped band depicts the sequence/chromosome.
- Narrowed areas on the rod indicate centromeric regions.
- Cytoband patterns are indicated by the vertical dark lines that run the length of the rod. Pause the cursor over a particular cytoband to display a tool tip with information about the cytoband.
- Centromeric regions and cytoband patterns are shown only for those DAS servers that provide the information in the correct format.

![Figure 33.11 Overview Pane](image)

**Note:** Cytoband data is downloaded from the DAS server. Currently (at the time of this publication), UCSC provides cytoband data for human chromosomes only.

A ruler, visible at the top of the Overview Pane, marks the length of the genomic sequence in 5 Mbp increments. The symbol indicates the fragment area currently shown in the Feature Map and Information Panes.

**Feature Map Pane**

The Feature Map Pane shows a graphical depiction of the features for the currently loaded fragment (Figure 33.12). You can filter the types of features you want to download for loaded fragments. See **Setting Feature Type Filters** on page 530.

![Figure 33.12 Feature Map Pane showing features for the loaded fragment](image)

Like the Overview Pane, the Feature Map Pane has a ruler at the top to help you visualize the orientation of features along the length of the fragment.
Feature types or tracks display along the left side of the Feature Map Pane. Different features are depicted with various graphical representations and are labeled with the feature name. Shown features are listed along with relevant feature information in the Features tab of the Information Pane. See Features Tab on page 532.

Features marked with an orange square (Figure 33.13) have sequences associated with them. The associated GenBank accession number is derived from the Target Accession number of the feature. If this number is a valid GenBank accession number, GenomBench can download the feature’s associated sequence (usually from NCBI) so the sequence can be used for alignments in the Alignment tab of the Information Pane. For more information, see Alignment Tab on page 523. You can download the features of such sequences, as well. URLs other than NCBI can be associated with any feature to download sequence from sources other than NCBI. For more information, see Specifying Data Source URLs for Feature Types with Alignment Information on page 537.

Note: Many of the operations discussed below for the Feature Map Pane are also available in the Features tab of the Information Pane.

### Feature Map Pane Viewing Options

The Feature Map Pane view can be changed in terms of feature layout, feature style, and/or ruler coordinate scale.

#### Feature Layout Options

In the Feature Map Pane, you can choose between two different feature layout types: Divided by Feature Types and Solid Layout.

- **Divided by Feature Types** – In this layout type, features are shown along the length of the fragment in groups according to type with the feature type indicated on the left side of the pane (Figure 33.10).

  In Divided by Feature Types layout, you can change the way in which features are ordered, displayed, and selected in the following ways.

  - **Change the feature listing order** by dragging and dropping the feature type names on the left side of the Feature Map Pane. The pane autoscrolls if there are many feature tracks displayed in the pane. If the Remember Tracks Order option is checked in the Graphic Style Editor Layout tab (see page 534), GenomBench remembers the new order and lists features in this order the next time a fragment is opened.

  - **Expand or collapse Feature types** either individually or globally by selecting Expand Type/Expand All Types or Collapse Type/Collapse All Types from the shortcut menu. To expand or collapse an individual type, the cursor must be in the feature type on the left side of the Feature Map Pane.

  - **Sort feature tracks alphabetically** by label by selecting Sort by Label from the shortcut menu.
Highlight all features in a type by selecting Select All Features in Track from the shortcut menu.

Edit the graphic style of features in a type by selecting Edit Graphic Style from the shortcut menu. This opens the Graphic Style Editor, which is described in Editing Graphic Styles on page 531, with the graphical properties of the selected feature type displayed.

- **Solid Layout** – In this layout type, features are shown along the length of the fragment where they occur and are not grouped according to feature type. There are no feature category labels at the left of the Feature Map Pane, but the feature type is shown in a tool tip when you pause the cursor over a feature.

To change the Feature Map Pane layout setting, select Layout Type > <layout type> from the shortcut menu or View > Layout Type > <layout type>. You can also click the Switch Layout button ( ) on the toolbar.

**Feature Style Options**

Two different feature style viewing options are available in the Feature Map Pane:

- **Full Feature Height** – Features are shown at full size (Figure 33.10).
- **Low Feature Height** – Features are condensed vertically to save space on the screen.

To change the feature style, select View > Feature Height > Full Feature Height or Low Feature Height. You can also click the Feature Height button ( ) on the toolbar.

**Coordinate Scale Options**

The Feature Map Pane ruler has two different coordinate scales:

- **Global Coordinates** – Ruler shows coordinates of the genomic sequence.
- **Fragment Coordinates** – Ruler shows coordinates of the current fragment sequence.

The coordinate scale is also reflected in the sequence numbering in the Sequence tab of the Information Pane.

To change the ruler scale, select View > Coordinates > Global Coordinates or Fragment Coordinates or right-click on the ruler and choose the corresponding commands from the ruler’s shortcut menu. You can also click the Switch Coordinates button ( ) on the toolbar.

**Creating, Editing, and Deleting Features**

In the Feature Map Pane, features can be created, edited, or deleted from commands available on the Feature menu and/or the Feature Map Pane shortcut menu.

- **Create New Feature** – Opens the Create New feature dialog box so you can create a new feature on the sequence fragment. The Create New Feature dialog box is almost identical to the Edit feature dialog box (see Figure 33.12 and Figure 33.13), except that the feature attribute values are unpopulated. The Reset to Selection and Reset to Mark(s) buttons set the feature position coordinates to the current selection or mark(s), respectively, in the Feature Map Pane/Sequence tab of the Information Pane.

- **Delete Selected Features** – Deletes the feature(s) currently selected in the Feature Map Pane/Features tab of the Information Pane.

- **Edit Feature** – Opens the Edit Feature dialog box where you can edit the current attributes for a feature.

Select a feature in the Feature Map Pane and select Edit Selected Feature from the shortcut or Edit menu to open the Edit Feature dialog box for that feature. The Edit Feature dialog box has three tabs: the Basic Attributes tab (Figure 33.14), the Additional Attributes tab (Figure 33.15) and the Links tab (Figure 33.16) in which you can edit the feature attributes. For features of
sequences that are opened from or have been saved to the Local database, a fourth tab displays, the Qualifiers tab (Figure 33.17).

In the Basic Attributes tab (Figure 33.14), you can view and edit basic parameters for the feature, such as feature name, type, orientation and component type and position on the fragment. Click the Create New Type button to define a new feature type. Click the Reset Segments button to reset the component section to the original values.

![Figure 33.14 Basic Attributes tab of the Edit Feature dialog box](image)

In the Additional Attributes tab (Figure 33.15), you can edit the listed attributes by highlighting the attribute, clicking the icon to the right of its value and typing in or selecting the new value from a list.

![Figure 33.15 Additional Attributes tab of the Edit Feature dialog box](image)

The Links tab is for associating a URL or a file directly to a feature (left side of Figure 33.16). Links are added by clicking the Add button in the Links tab. In the Add Feature Link dialog box that opens (right side of Figure 33.16), either enter or paste a URL or the location of a file in the Link URL field. You can also click the Browse button to locate a file. Enter a descriptive label in the Link Label field; this label is displayed in the shortcut menu of the Feature Map Pane when the feature is selected. Edit a link by selecting the row and either entering new information directly into each field or clicking the Edit button, which opens the Edit Feature Link dialog box. Remove a link by selecting the link and clicking the Remove button.

To open a link, click the Link value to open the file in the appropriate program or the URL to the appropriate web page. Links can also be added by dragging and dropping files or URLs directly onto a feature in the Feature Map Pane. For more information, see Adding Links to Features on
Links are only saved with a feature when the fragment is saved to the local database. For more information, see Saving Sequences on page 552.

For features of sequences that have been saved to the local GenomBench database, a fourth tab, the Qualifiers tab (Figure 33.17), is present. For sequences with associated GenBank entries (features marked with an orange square; see Figure 33.13 and related text), the Qualifiers tab can list any qualifiers that are present in the GenBank entry. To populate the Qualifiers tab with information from an associated GenBank entry, you must transfer the qualifier information from the GenBank entry of the target sequence using the **Transfer Features of Target Sequence** command in the Advanced shortcut menu to create a component feature containing the information. See **Target Sequences** on page 521.

**Note:** As described above (see Figure 33.13 and related text), the associated GenBank accession number must be valid before GenomBench can download the feature’s associated qualifiers from NCBI.

Opening the Edit Feature Source dialog box for the newly created component, shows the qualifier information from the GenBank entry for that feature (Figure 33.17):

You can edit the values for listed qualifiers by highlighting the qualifier, clicking the icon to the right of its value and entering or selecting the new value from a list. To add new qualifiers, click the **Add** button and configure the new qualifier in the same way that you edit existing qualifiers.
Note: You can view a feature’s qualifiers in the Qualifiers column of the Features tab. If the Qualifiers field is not displayed, add it to the Features tab display using the instructions in Customizing the Columns View on page 527.

When you have finished editing the feature properties, click OK to submit your changes.

Editing Multiple Features Simultaneously

Multiple features can be added or edited simultaneously by first selecting a feature, selecting one or more additional features by pressing the CTRL key, and then opening the Edit Features dialog box. For example, you may want to add the same URL or change the type on multiple features. Fields that have more than one value (for example, the Name field) display '<multiple values/no change>'. Changing the value of these fields will reset the value in all selected features.

Feature Grouping/Ungrouping Operations

Feature grouping operations allow you to create composite features and move features in and out of them. Several feature grouping operations are available on the Feature Map Pane shortcut menu.

Creating composite features provides a convenient method for grouping related features, such as the promoter, UTRs, CDSs, exons, etc. for a gene. In this way, you can keep all related features together, so you can view and work with them in close proximity in the Feature Map Pane and Features tab. Composite features can also be used to contain the features of a larger feature.

Grouping operations are described as follows:

- **Group Selected Features** – Allows you to group the selected features into one composite feature. To group features, select Advanced > Group Selected Features from the shortcut menu. This launches a dialog box similar to the Edit Feature dialog box (see Feature Properties Dialog Box on page 520) where you can define attributes for the new composite feature. Click the +/− symbol at the left of the composite feature to expand it and show the features that compose it or contract it to show only the composite feature.

- **Move Selected Features Into** – Moves the selected feature(s) into a group. This is accomplished by dragging and dropping the feature(s) onto the group or feature.

- **Move Selected Features Out** – Moves the selected feature(s) out of their group. This is accomplished by dragging and dropping the feature(s) out of their composite group.

Feature Properties Dialog Box

The Feature Properties dialog box allows you to view the attributes of the selected feature. The feature properties dialog box is identical to the Edit Properties dialog box (see Creating, Editing, and Deleting Features on page 517), except that it is designed for viewing properties, not editing them.

To open the Feature Properties dialog box for a feature, highlight the feature in the Feature Map Pane or Features tab of the Information Pane and select Feature > Properties or Properties from the shortcut menu.

Adding Features to an Alignment

Features with associated sequence (see Figure 33.13) can be downloaded from NCBI or other sites and aligned against the genomic sequence in the Alignment tab of the Information Pane. Features that have associated sequence are indicated by an orange box in the Feature Map Pane. To add feature(s) to the Alignment tab, select one or more features in the Feature Map Pane or Features tab of the Information Pane by pressing the CTRL key and clicking on each feature or pressing the CTRL key and dragging a rectangle around the features to be down-
loaded. Select Feature > Add Features to Alignment or click and drag them from the Feature Map Pane onto the Alignment Tab.

While the features are downloading, the right  icon in the lower right of the status bar becomes active. If the process appears to be time-consuming, you can abort the download by clicking on the icon. The Data Loading Monitor opens. Click on the Supporting Information tab, select one or more sequences, and click the Abort button to cancel the process.

**Target Sequences**

You can perform the following operations on features that have target sequences associated with them (for example, GenBank entries that are downloadable):

- **Open Target Sequence** – Opens the target sequence for the selected feature in the current GenomBench window, replacing the current sequence. Select Advanced > Open Target Sequence from the shortcut menu.

- **Open Target Sequence in New Window** – Opens the target sequence for the selected feature in a new GenomBench window. Select Advanced > Open Target Sequence in New Window from the shortcut menu or click the button on the toolbar.

- **Transfer Features of Target Sequence** – Transfers the features of the target sequence for the selected feature into the current GenomBench window. Select Advanced > Transfer Features of Target Sequence from the shortcut menu or click the button on the toolbar.

**Important:** You must save the fragment you are viewing to the local database before the Target Sequence commands become enabled.

**Note:** The States column in the Features tab of the Information Pane contains symbols that indicate whether or not a particular feature has an associated target sequence. For more information on the States column, see Symbols in Features Tab Fields on page 526.

**Adding Links to Features**

Files and URLs can be linked directly to features in the Feature Map Pane by dragging and dropping. Links can also be added in the Link Tab of the Edit Feature dialog box, described in Creating, Editing, and Deleting Features on page 517. To link a file or URL to a feature, drag and drop the file on the local file system or the browser symbol (for example, for Internet Explorer) from the title bar or the address bar of the browser onto the feature in the Feature Map Pane. The cursor changes from a symbol to an arrow indicating that you can drop the file or URL onto the feature. Click OK to confirm that you want to add the link to the feature.

Each link added to a feature displays when Links is selected from the shortcut menu. Selecting a link from the shortcut menu opens the file in the appropriate program or the URL to the appropriate web page. Added links also display in the Links tab of the Properties and Edit Feature dialog boxes. Links are only saved with a feature when the fragment is saved to the local database. For more information, see Saving Sequences on page 552.

**Information Pane**

The Information Pane has four tabs, Sequence, Alignment, Features and Text, which display different types of information about the genomic sequence or current fragment sequence displayed in the Feature Map Pane.

**Note:** An additional tab, the Search Results tab, is added to the Information Pane when a Feature Search is performed. See Feature Search Tool on page 539.
Sequence Tab

The Sequence tab of the Information Pane displays the sequence of the current fragment shown in the Feature Map Pane (Figure 33.19). In the Sequence tab, you can perform basic operations such as copying and marking/unmarking selected parts of the sequence. You can also show/hide features selected in the Feature Map Pane, edit the sequence, and perform translations.

Note: The coordinate scale at the left of the sequence can be set to show global (genomic sequence) or fragment coordinates. See Coordinate Scale Options on page 517.

Copying Sequences

Sequences can be copied directly from the Sequence tab to the clipboard and pasted into word processing programs. To copy a sequence, select the sequence and select Copy from the shortcut menu or click the Copy button on the toolbar. This copies the direct strand. To copy the complementary strand, select Copy Complementary from the shortcut menu.

Showing/Hiding Features in the Sequence Tab

To show a graphic representation of a feature(s) in the Sequence tab, select the feature(s) in the Feature Map Pane or the Features tab, right-click on the Sequence tab, and select Show Selected Features from the shortcut menu. The feature is marked with a red line and the feature name in the Sequence tab above the area of the sequence where the feature(s) occurs (Figure 33.19). More than one feature can be shown in the Sequence tab at a time.

To hide features in the Sequence tab, select Hide All Features from the Sequence tab shortcut menu.

Figure 33.18  Sequence Tab of the Information Pane

Figure 33.19  Displaying a highlighted feature in the Feature Map Pane on the Sequence tab

Figure 33.19 Displaying a highlighted feature in the Feature Map Pane on the Sequence tab

To hide features in the Sequence tab, select Hide All Features from the Sequence tab shortcut menu.
Translating Sequence Selections

Selected sequence in the Sequence tab can be translated directly. To translate a portion of the sequence, highlight it in the Sequence tab and select Translate from the shortcut menu. The translation appears above the selected sequence in the Sequence tab (Figure 33.20):

![Figure 33.20 Selection translated in Sequence tab](image)

You can select two or more non-contiguous regions, marking each one after its selection with the Mark Selected Segments button ( ). Then right click in the Sequence tab and select Translate Mark > Direct or Complementary. This translation feature enables you to translate non-contiguous DNA blocks (such as exons) in the correct frame.

To translate a different region or the same region in a different frame, make the new selection and repeat the process.

To remove all translations in the Sequence tab, select Clear All Translations from the shortcut menu.

Alignment Tab

Features marked with an orange square in the Feature Map Pane (Figure 33.12) have genomic sequence alignment data, also called target sequence, associated with them. The sequences from such features can be aligned against the genomic sequence in the Alignment tab.

To add feature sequences to the Alignment tab of the Information Pane, highlight the feature(s) in the Feature Map Pane or Features tab of the Information Pane and select Feature > Add Features to Alignment. Features can also be added to the Alignment tab by dragging them from the Feature Map Pane onto the Alignment tab.

Feature sequences appear in the Alignment tab aligned against the genomic sequence (Figure 33.21). Placing the cursor over any base in the alignment displays a tooltip with information about the position on the genomic sequence, the position in the alignment, and the cumulative number of each type of base at that position.

![Figure 33.21 Feature sequences aligned with genomic sequence in Alignment tab](image)

A number of operations are available from the buttons at the top of the Alignment tab and from the Alignment tab and ruler shortcut menus. These options are discussed in the following sections.
Alignment Color Scheme

Residues in the alignment are colored according to the following scheme:

- **Black on window default color** – Non-similar to other residues at that position.
- **Blue on Cyan** – Residue is similar to other residues at that position.
- **Black on Green** – Residue occurs greater than 50% of the time at that position.
- **Red on Yellow** – Residue is completely conserved at that position.
- **Green on window default color** – Residue is weakly similar to the consensus residue at that position.

Alignment Tab Viewing Options

In the Alignment tab, a coordinate scale ruler appears above the alignment. The scale numbering can be based on the genomic sequence (global) or on the fragment coordinates, depending on the viewing mode you have selected. See *Coordinate Scale Options* on page 517.

You can also set the scale numbering to reflect the alignment coordinates. To do this, right-click on the ruler in the Alignment tab and select *Show Alignment Coordinates* from the ruler shortcut menu. The coordinates now begin with position 1 at the beginning of the alignment.

Alignment Tab Navigation Operations

The following navigation commands are available from the Alignment tab tool buttons and by choosing *Navigation > Go To < command>* from the shortcut menu:

- **Go To Left Edge** (←) – Moves to left edge of the current sequence.
- **Go To Right Edge** (→) – Moves to right edge of the current sequence.
- **Go To Alignment Start** (↑) – Moves to the beginning of the alignment.
- **Go To Alignment End** (↓) – Moves to the end of the alignment.

**Note:** An “edge” can be the end of the sequence or the transition point from gap to no gap (and vice versa) in the sequence.

Other keyboard shortcuts include:

- **ARROWS** – Move the cursor one position in the direction of the arrow.
- **PAGE UP, PAGE DOWN** – Move the cursor to first or last sequence, respectively.
- **CTRL+OPERATION** – Applies focus rather than selection.
- **SPACE** – Applies the selection to the area in focus.
- **CTRL+A** – Selects all sequences.
- **SHIFT+ARROW** – Selects a range or extends a selection, one position at a time.
- **BACKSPACE** – Deselects a selection.
- **CTRL+SHIFT+ARROW** – Extends current selection to the end of the sequence indicated by the arrow.

In addition, double-clicking with the mouse selects the segment (exon) or gap (intron) and shifts the selection to that sequence. Selections can also be made by clicking and dragging with the mouse.
Alignment Tab Editing Operations

The following general editing operations are available from the Alignment tab tool buttons and shortcut menu:

- **Edit Alignment** ( ) – Switches Alignment tab to Edit Mode, activating all other editing operations.

**Important:** You must switch to Edit Mode to enable all edit commands except the **Remove Features from Alignment** command.

- **Submit Changes** ( ) – Registers the editing changes you have made.
- **Cancel Changes** ( ) – Cancels the editing changes you have made.

The following commands for moving selections are available from the Alignment tab tool buttons and by selecting **Edit Selection > Move <command>** from the Alignment tab shortcut menu:

- **Move 1 Left** ( ) – For a selected portion of a sequence with a gap to its left, moves the selected sequence to the left one residue.
- **Move 10 Left** ( ) – For a selected portion of a sequence with a gap of at least ten residues to its left, moves the selected sequence to the left ten residues.
- **Move Far Left** ( ) – For a selected portion of a sequence with a gap to its left, moves the selected sequence to the left, closing the gap.
- **Move 1 Right** ( ) – For a selected portion of a sequence with a gap to its right, moves the selected sequence to the right one residue.
- **Move 10 Right** ( ) – For a selected portion of a sequence with a gap of at least ten residues to its right, moves the selected sequence to the right ten residues.
- **Move Far Right** ( ) – For a selected portion of a sequence with a gap to its right, moves the selected sequence to the right, closing the gap.

You can also drag and drop a block of text to an adjacent gap by selecting the block and pressing the CTRL-key while dragging.

The following commands for inserting and deleting gaps are available from the Alignment tab shortcut menu by selecting **Edit Gaps > <insert/delete gap command>** or by clicking the corresponding button on the Alignment tab toolbar:

- **Insert Gap** ( ) – Inserts a gap in the selected feature the size of the highlighted sequence, moving the sequence to the right, beginning at the first highlighted residue.
  
  When you select all the molecules in an alignment including the genomic sequence, you can introduce gaps in the whole alignment that are represented as dashes.
- **Delete Gap** ( ) – Deletes the highlighted gap in the selected feature.
- **Remove Gap-only Columns** ( ) – Deletes gaps in entire alignment that are introduced when all molecules including genomic sequence are selected and a gap is inserted.

You can also resize a gap by selecting the boundaries of the gap and then, with the cursor over one of the boundaries, press the CTRL-key and drag the boundary to the new position.

When you introduce gaps in an alignment, the genomic sequence is extended on the far right of the alignment view. However, if you delete the gaps, GenomBench does not automatically remove the added sequence at the far right. If you want to see only the range that is actually used by the aligned molecules, select **Truncate Alignment** from the shortcut menu or click the button on the Alignment tab toolbar.
To remove a feature from the alignment, select the feature name in the Alignment tab and select **Remove Features from Alignment** from the shortcut menu.

**Features Tab**

The Features tab of the Information Pane lists the features shown for the current fragment and information related to those features in a table (Figure 33.22). Feature information is organized into various fields in the Features tab.

![Figure 33.22 Features tab of the Information Pane](image)

**Symbols in Features Tab Fields**

Two of the fields in the Features tab, **States** and the field just to the right of the States field, usually the **Name** field, contain symbols with the following significance:

- There are four symbols in the **States** field, which indicate the following conditions about the sequence:
  - ♦ ✧ – Indicates whether the feature has sequence data, also known as target sequence, associated with it (for example, a GenBank entry), available from a server.
  - ✧ – Indicates whether the feature is currently displayed in the Sequence tab.
  - ✧ – Indicates whether or not the feature has genomic alignment data associated with it, available from the server.
  - ✧ – Indicates whether the feature is shown in the current alignment in the Alignment tab.

  An available symbol indicates that the condition is true; a greyed-out symbol indicates that the condition is false.

- When the field just to the right of the States field has a triangle next to the data in the field, this indicates that the data is in hierarchical order, as it was received by the server (that is, it has not been sorted by column).

**Sorting Features in the Features Tab**

Initially, features come into GenomBench in the order in which they are received by the server or hierarchical order. You can sort features according to any of the displayed fields by clicking the column header. Clicking a column header sorts the features in ascending order according to that field. Clicking the same column header a second time sorts the features again, this time in descending order according to that field. Clicking a column header a third time returns the results to hierarchical order. The symbol next to the column header name indicates the sorting status:

- [ ] – Features sorted in original or hierarchical order.
- ✧ – Features sorted in ascending order according to this column.
- ✧ – Features sorted in descending order according to this column.
Customizing the Columns View

You can customize the columns shown in the Features tab in the Customize List Columns dialog box. To launch this dialog box (Figure 33.23), select Customize from the Features tab shortcut menu. By default, the columns Name, Type, Start, Length, Orientation, and Category display in the Shown Columns list on the right side of the dialog box and subsequently display in the Features tab. All other columns that can be displayed in the Features tab are in the Available Columns list on the left side.

![Figure 33.23 Customize List Columns dialog box](image)

To remove a column from view, highlight it in the Shown Columns list on the right and click the Hide button to move it to the Available Columns list. To add a column back to the view, highlight it in the Available Columns list and click the Show button to move it into the Shown Columns list.

Change the column order in the Shown Columns list by selecting a column name and clicking the Move up or Move down buttons. Restore the defaults by clicking the Restore Default button.

You can also show columns and hide columns directly from the Features tab. To hide a column, put the cursor on the column header and select Hide column from the shortcut menu. To show a column, put the cursor on any column header, select Show column, and select the column from the submenu.

Features Tab Operations

A number of other operations are available on the Features tab shortcut menu, including general feature operations, feature grouping and displaying feature properties. These operations are discussed in the Feature Map Pane section on page 515.

Highlighting a feature in the Features tab also highlights the feature in the Feature Map Pane. Double-clicking a feature in the Features tab opens the Properties dialog box for the feature (see page 515).

Double-clicking in a cell in the Feature tab allows you to perform in-cell editing where available or opens the Edit Features dialog box. When in-cell editing is available, you can directly change the value. To change a value in the Edit Features dialog box, see Creating, Editing, and Deleting Features on page 517.
Text Tab

The Text tab of the Information Pane (Figure 33.24) shows the attributes for the sequence shown in the Overview Pane. Basic data about the sequence, such as ID, name, length, download path and current fragment coordinates can be viewed here. In addition, when alignment analyses are performed with Spidey or Sim4, their results appear in the Text tab. See Sim4 and Spidey Alignment Tools on page 542.

![Figure 33.24 Text tab of the Information Pane](image)

Notes:
- If you have not saved the fragment to your local database, some of the above information does not appear in the Text tab.
- Clicking on the Original URL hyperlink in the Text tab opens the original sequence from the server in a new GenomBench window with the current fragment information displayed.

Editing Sequence Attributes

To edit sequence data in the Text tab, select Edit > Edit Sequence Attributes from the GenomBench menu or select Edit Sequence Attributes from the shortcut menu. The Edit Sequence Attributes dialog box displays and has two tabs: Basic Attributes (Figure 33.25) and Additional Attributes (Figure 33.26). In the Basic Attributes tab, you can edit the name and description of the sequence.

![Figure 33.25 Basic Attributes tab of the Edit Sequence Attributes dialog box](image)
In the Additional Attributes tab, you can edit the version number, accession number, URL to the original sequence, etc. When you have completed your edits in the Edit Sequence Attributes dialog box, click OK to accept the changes. The Text tab is updated with the new information.

![Figure 33.26 Additional Attributes of the Edit Sequence dialog box](image)

**Undo and Redo Operations**

A number of the operations that you can do in the Feature Map Pane, the Feature tab, and the Alignment tab can be undone or redone. A history is also kept of each undone or redone operation so you can easily make changes even after you have made several subsequent changes.

Undo and redo can be used on the following operations as they are performed in the Feature Map Pane and the Feature and Alignment tabs of the Information Pane:

- Adding, editing and deleting features
- Converting temporary features (analyses) into permanent features
- Showing and hiding feature types
- Adding and removing features in alignments
- Alignment editing including:
  - moving blocks
  - resizing gaps
  - moving residues to the left or right
  - inserting and deleting gaps
- Submitting and cancelling changes

The undo and redo commands are initially unavailable until one of the above operations is performed. Once an operation is performed, the Edit > Undo command and the Undo button on the toolbar are available. The undo command reverts the previous change and places it in the redo list. Once an undo operation is performed, the Edit > Redo command and the Redo button on the toolbar are available. The redo command redoes the last undone change and places it back in the undo list.

When a change is made, a record of the change is added to a Undo/Redo History log. To open the History log, select Edit > View History. The current state of the GenomBench window is designated by a green arrow in the History log. Any changes are listed above the current state and can be selected and undone as in the left side of Figure 33.27 by clicking the Undo button.
When a change is undone, it is moved below the current state as in the right side of Figure 33.27 and can be redone by clicking the **Redo** button.

![History Log](image)

*Figure 33.27 History log displaying changes that can be undone and redone above and below Current state*

Groups of operations can also be undone or redone simultaneously. Selecting an operation that is listed more than one away from the current state selects all operations in between. Clicking the Undo or Redo button either undoes or redoes the selected operations as a group and moves them into the redo or undo list respectively.

**Configuring GenomBench Parameters**

Many of the parameters relating to fragment loading and display in GenomBench can be customized. A number of customization commands are accessible individually under the **View** menu and on the Feature Map Pane shortcut menu. These commands are described in *Feature Map Pane Viewing Options* on page 516.

GenomBench also offers several convenient dialog boxes where the complete list of customization options for feature downloading and graphic styles display are consolidated for easy editing.

**Setting Feature Type Filters**

Feature types that you want to download from the current server for the current sequence are specified in the Display Feature Types dialog box (Figure 33.28).

**Note:** The Display Feature Types dialog box shows only those features that can be downloaded from the current server; it does not show global feature types (that is, features available from all DAS servers). For information about setting global feature types options, see *Editing Fragment Loading Configurations* on page 534.
To launch the Display Feature Types dialog box, select View > Display Filters or click the Feature Type Filter button ( ) on the toolbar.

![Feature Types Filter dialog box](image)

In the Display Feature Types dialog box, uncheck the feature types that you do not want to download (by default, all feature types are selected for download). The Select All and Clear All buttons allow you to check or uncheck, respectively, all the feature types in the list at once. Click OK to apply the changes. Now, when you download a fragment, only the checked feature types are downloaded.

**Important:** Once a fragment is saved to Local storage, only those features that are downloaded are saved with the fragment. You cannot then reopen the fragment from Local storage and see more features types by removing feature type filter constraints. In this case, if you are not sure what types of features you may be interested in for this fragment, it is recommended that you download all feature types so you can save the fragment to local storage and open it later displaying all feature types.

**Editing Graphic Styles**

GenomBench’s Graphic Style Editor allows you to specify display parameters for the Feature Map Pane, including Feature Map Pane layout characteristics, fragment line, pattern and color characteristics, and feature line, pattern, color and shape characteristics.

To launch the Graphic Style Editor, select View > Edit Graphic Styles. The Graphic Style Editor has three tabs: Features, Fragment and Layout. It also has Save As and Load buttons for saving your custom graphics configurations to files and loading graphics configurations from previously saved files, respectively.
**Features Tab**

The Features tab (Figure 33.29) contains parameters that allow you to determine the way different features appear in the Feature Map Pane.

![Figure 33.29 Features tab of the Graphics Style Editor dialog box](image)

To view or change the current style for a feature, highlight the feature name in the Feature Types list. Set the line thickness/style/color, pattern style/color, and/or shape. A preview of the new configuration displays in the Preview panel on the right.

Click the **Predefined Settings** button to return all feature styles to their default settings.

Click **OK** to apply your changes.
**Fragment Tab**

The Fragment tab (Figure 33.30) contains parameters that allow you to determine the way the graphical depiction of the fragment appears in the Feature Map Pane.

Figure 33.30 Fragment tab of the Graphics Style Editor dialog box

To change the display style for the fragment, select the line thickness/style/color and/or pattern style/color. A preview of the new configuration displays in the Preview panel on the right.

Click the **Predefined Settings** button to return the fragment display to default settings. Click **OK** to apply your changes.
**Layout Tab**

The Layout tab (Figure 33.31) contains parameters that allow you to determine styles and colors relating to Feature Map Pane layout and marking. Multi-segmented feature style is also set here.

![Figure 33.31 Layout tab of the Graphics Style Editor dialog box](image)

In the **Feature** section, set the values in pixels for Low and Full Feature Height layouts, select the font typeface and color for feature labels in the Feature Map Pane and choose the style for Multi-segmented Features.

In the **Marks** section, set the color and pattern for marks in the Feature Map Pane and Sequence tab of the Information Pane.

In the **Types** section, set the background color and font typeface and color for labeling in the feature category section of the Feature Map Pane.

You can also set the pane divider, connector and segment selection color in the Layout tab.

In the **Feature Map Margins** field, set the value in pixels between 0 and 20 to change the size of the margin on the right and left sides of the area where feature tracks display in the Feature Map Pane.

Check the **Remember Tracks Order** if you want GenomBench to remember the feature tracks order (feature type order when in Divided by Feature Types layout mode). When you open GenomBench again, it will list feature types in this order. Otherwise, feature tracks are listed in alphabetical order according to feature name.

Click the **Predefined Settings** button if you want to return the layout display to default settings. Click **OK** to register your changes.

**Editing Fragment Loading Configurations**

In the Configurations dialog box, you can set parameters for fragment/sequence downloading size warnings, select the feature types you want to globally download, and associate a URL with sequence that has alignment information. If a fragment exceeds the set download size, you have the option to download sequence on demand. For more information, see Downloading Sequence on Demand on page 535.
To launch the Configurations dialog box, select View > Configuration. The Configurations dialog box has three tabs: Warning Triggers (Figure 33.32), Cumulative List of Feature Types (Figure 33.34), and Warning Messages (Figure 33.38).

**Warning Triggers Tab**

The Warning Triggers Tab allows you to set parameters to trigger warning messages if downloaded fragments and sequences fall outside of set ranges.

- **Limits for fragment size to be downloaded** – Sets the maximum size fragment to download. If you try to download fragments larger than the warning size, a warning message displays. You can either truncate the fragment or download in Sequence on Demand mode.

- **Warning limits for sequence size to be sent to Vector** – Sets the maximum size for sending a sequence to Vector NTI and the maximum size for sending a sequence to Analysis Monitor for Sim4 and Spidey alignment analysis. For more information on sending sequences to Vector NTI and Analysis Monitor, see GenomBench Tools on page 539.

- **Size of flanking region to be added to the fragment by Query Application** – When GenomBench opens a region of a chromosome found by the Query Application, this parameter determines how many flanking base pairs GenomBench should download as specified in the textbox.

**Downloading Sequence on Demand**

If you try to download a fragment larger than the **Limits for fragment size to be downloaded** field in the Warning Triggers tab, you will be given the options to either truncate the fragment to the beginning segment of the pre-defined size or download it in a mode called ‘Sequence on Demand’. In Sequence on Demand mode, GenomBench initially downloads only the features of the selected fragment but not the sequence. Sequence is then retrieved as needed based on operations performed in the GenomBench window. For example, when a particular feature such as an EST is selected, the sequence is downloaded and shown in the Sequence tab.

As an example of how Sequence on Demand mode works, consider the case where the maximum fragment size to download is 2Mbp. If you were to select a fragment of 3Mbp, the dialog box in Figure 33.33 displays. Select the top radio button to load the features and sequence for
the first 2 Mbp of the fragment. Select the lower radio button to load in Sequence on Demand mode and load the features that span all 3Mbp without sequence.

![Load large Fragment dialog box]

Figure 33.33 Load large fragment dialog box

As the features load in the Feature Map Pane, question marks initially display in the Sequence tab and as you scroll through it, they are replaced by the sequence in that region of the fragment. Likewise, as you select features in the Feature Map Pane and show them in the Sequence tab or drop them into the Alignment tab, sequence is loaded as it is requested.

Partially loaded fragments cannot be saved to the local GenomBench database. When saving a fragment that is loaded in Sequence on Demand mode, you will be prompted whether you want to load the whole sequence before saving. Click Yes in the dialog box that displays to save the sequence to the local database. It may take some time for the entire sequence to load. For more information, see Saving Sequences on page 552.

**Cumulative List of Feature Types Tab**

The Cumulative List of Feature Types tab allows you to globally set the types of features you want to download when loading a fragment (Figure 33.34). The feature list on this tab shows feature types from all DAS servers you use and is automatically updated when you access a DAS server.

![Cumulative List of Feature Types tab of the Configuration dialog box]

Figure 33.34 Cumulative List of Feature Types tab of the Configuration dialog box

In the Cumulative List of Feature Types dialog box, uncheck the feature types that you do not want to download (by default, all feature types are selected for download). The **Select All** and **Clear All** buttons allow you to check or uncheck, respectively, all the feature types in the list at once. Click **OK** to register your changes. For any subsequent sequence opened from a DAS server, only the checked feature types are downloaded for a fragment.
Note: If you have downloaded a fragment with some features unchecked in the Cumulative List of Feature Types tab (or in the Display Feature Types dialog box – Figure 33.28) and if you are still connected to the DAS server, you can then check the unchecked features in the dialog box and they will then download from the server. However, since these features are downloaded as a secondary step, they are listed in the Feature Map Pane as the last entry at the top of the list; they do not appear in alphabetical order within the first set of downloaded features.

Specifying Data Source URLs for Feature Types with Alignment Information

When a DAS server provides alignment information, you can specify a data source URL for the feature type. The specified URL directs GenomBench to a location to download the target sequence. Once a valid URL has been associated with a feature, features can be analyzed in the Alignment tab, the target sequence can be opened, and the features can be transferred from the target sequence. For more information, see Target Sequences on page 521.

URLs are associated with features from the Cumulative List of Feature Types tab. Select a feature type from the list in the tab and click the Edit URL button. In the Edit Feature Type Loading URL dialog box, enter the URL to be associated with that feature type in the URL field and click OK (Figure 33.35).

The URL for that feature type now displays in the Reference URL field in the tab (Figure 33.36). Click OK.

In the Feature Map Pane, the feature now contains an orange rectangle indicating it has target sequence associated with it. The circled feature (blastzBestMm3) in the following figure con-
tains mouse alignment against human sequence using BLAST. When dragged into the Alignment tab, the mouse sequence is shown aligned to the human sequence.

**Warning Messages Tab**

The Warning Messages tab (Figure 33.38) allows you to control the display warning dialogs that remind you what to do in specific situations. For example, if you are in Sequence on Demand mode and you want to save the sequence, you will be warned with a message that you can only do so if you download the entire sequence. Unchecking the **Saving partially loaded sequence** message, in this case, causes the message to not display. Check the messages you want to see; uncheck the messages you do not. Click the **Clear All** button to uncheck all messages; click the **Select All** button to select all messages.
GenomBench Tools

GenomBench includes Feature Search and Pattern Search tools for searching for features and sequences among those downloaded that meet criteria that you set. It also provides tools for aligning sequences to a genomic backbone using Sim4 or Spidey analyses and sending a fragment to Vector NTI. In addition, GenomBench provides a convenient interface for querying public databases.

Feature Search Tool

The Feature Search tool in GenomBench allows you to search for features from among the features downloaded for a fragment according to criteria that you define.

To launch a feature search, select Tools > Feature Search. The Feature Search Parameters dialog box displays (Figure 33.39):

![Feature Search Parameters dialog box](image)

In the Feature Search Parameters dialog box, define the **Search region** in base pair coordinates. If you make a selection in the Feature Map Pane before you launch this dialog box, the coordinates of the selection are pre-entered into the Search region fields.

Define your first search criterion according to **Field Name**, which corresponds to the columns in the Features tab of the Information Pane, and Condition using the drop-down lists for each field. Highlight the Value field and enter the text or value for that search criterion. An example of a search criterion is ‘Type Equals acembly’.

If you have another search criterion, select the **AND** or **OR** radio button, depending on whether you want to search for both criteria or either criteria, respectively, and click the **Add** button. A new entry appears in the search condition definition list. Define the new criterion as described above. If you want to remove a search condition, highlight it and click the **Remove** button.

Notice that the Query String at the bottom of the dialog box is automatically entered as you define your search criteria in the search condition definition list.
When you have finished defining your search criteria, click the **Start** button to initiate the feature search. When the search completes, the results display in a new tab, the Search Results tab, which is added to the Information Pane (Figure 33.40):

![Figure 33.40 Search Results tab added to Information Pane](image)

If additional feature searches are performed, new Search Results tabs are added to the Information Pane and are numbered consecutively in the order in which the searches are performed.

The same operations that are available for features on the Features tab are also available for features on the Search Results tab(s). See *Features Tab Operations* on page 527.

**Pattern Search Tool**

The Pattern Search tool in GenomBench allows you to search for stretches of sequence that match a defined entry. You can either enter a sequence or a regular expression as a search query. A regular expression describes a pattern or sequence of adjacent characters using a specific syntax. The equivalent of a wildcard in regular expression syntax is described here. Additional regular expression syntax is described in Appendix F.
Pattern searches are performed from the Pattern Search dialog box, which is opened by selecting **Tools > Pattern Search** (Figure 33.41).

In the field at the top of the dialog box, enter the pattern to search for. You can paste in a pattern from the clipboard or enter one. To enter a wildcard, use the character `.`. For example, the pattern `.` searches for any three-base pattern ending in TT.

**Note:** In a regular expression pattern, spaces are treated as characters.

When searching for a pattern, you can search the whole molecule, a selected region, a specified region where you enter start and end coordinates, or marked regions. In the **Search Region** box, select the radio button that corresponds to the region you want to search. To search a selected region, highlight a region in the Sequence Pane or in the Feature Map Pane.

In the **Strands to Search** box, select the radio button corresponding to Direct, Complementary or Both.

Results are displayed either in the form of marks or temporary features. In the **Results** box, select the way you want the results to be displayed. You can create up to 1000 features from a single search. Temporary features must be converted to permanent features in order to be saved with a fragment. For more information, see Converting Temporary Features to Permanent Features on page 543. Marks are described in Marking and Unmarking Selections on page 514.

Click **Start**. When the search is complete, the progress bar is filled in. If features were selected as the results display type, the number of hits displays also. Click **OK** to show the results (Figure 33.42). Marked patterns display as blocks in the Sequence Tab. To move from one mark to the next, select **Edit > Select Marked Segment > Next/Previous** or press SHIFT+> or SHIFT+<. Temporary features display in a newly added feature track called Pattern Search Result as a continuous block. To view the individual features, click the + sign next to the block. Temporary features must be converted to permanent features in order to be saved with a fragment. For more information, see Converting Temporary Features to Permanent Features on page 543.
Sim4 and Spidey Alignment Tools

GenomBench provides the ability to analyze a fragment using the algorithms Sim4 and Spidey. Sim4 is a similarity-based tool for aligning an expressed DNA sequence (EST, cDNA, mRNA) with a genomic sequence. For more information about Sim4, see http://www.hgmp.mrc.ac.uk/. Spidey performs a similar function aligning an mRNA to a genomic sequence. For more information about Spidey, see http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html.

Before you attempt to perform a Sim4 or Spidey analysis, the following conditions must be met:

- The fragment to be analyzed must be saved to the local GenomBench database before a Sim4 or Spidey analysis can be performed on it.
- The cDNAs, ESTs, and mRNAs to be used for the analysis must be stored in their own unique subbase in the Vector NTI local database before they can be used in the analysis.

Performing a Sim4 or Spidey Alignment

To launch a Sim4 or Spidey analysis, first save the fragment to the local GenomBench database (if not already saved) and select Tools > Align cDNA with Genomic DNA using Sim4 or Align cDNA with Genomic DNA using Spidey, respectively.

GenomBench’s Sim4 and Spidey analyses are performed in the Analysis Monitor. When you execute a Sim4 or Spidey analysis, the Analysis Monitor opens with the appropriate interface and information for the fragment you are analyzing. For more information about using Analysis Monitor for Sim4 and Spidey analyses, see page 562 and page 564, respectively of Chapter 34, Analysis Monitor - Advanced Analysis Tools.

Viewing Sim4 and Spidey Results

Once a Sim4 or Spidey analysis has been performed in Analysis Monitor, you can send it back to GenomBench to review the results of the analysis. See Sending Analyses to GenomBench on page 571. The alignment analysis information displays in the Text tab of the Information Pane and the hits display in the Feature Map Pane and Features tab of the Information Pane (Figure 33.43 and Figure 33.44).

You can also send the analysis results to Vector NTI (see Sending Analyses to Vector NTI on page 571, page 571), but often this may not be practical due to the large size of most genomic fragments.
When a Sim4 or Spidey analysis is opened in GenomBench, alignment hits come in as temporary features and are placed in the Hit feature category in the Feature Map Pane and Features tab of the Information Pane (Figure 33.43 and Figure 33.44).

**Important:**
- Temporary features cannot be saved. They must be converted to permanent features if you want to save them with the sequence to the local database. If hits are not converted to permanent features, they are lost without warning upon saving.
- Analysis information shown in the Text tab cannot be saved with the sequence. However, the original analysis is stored in the Analysis Monitor window (it can also be saved to a file from Analysis Monitor), so you can re-launch the analysis in GenomBench from the Analysis Monitor window if you want to review this information. For more information, see *Saving* on page 571 and *Sending Analyses to GenomBench* on page 571.

**Converting Temporary Features to Permanent Features**

In GenomBench, you can convert hits or temporary features, such as those displayed as a result of a Spidey or Sim4 alignment search, to permanent features.
Hits are converted to permanent features by selecting the feature(s) and then selecting **Advanced > Transform Selected Features to Permanent** from the shortcut menu in the Feature Map Pane (Figure 33.44).

![Figure 33.44 Sim4 hits shown in Feature Map Pane and Features tab](image)

The feature is converted from temporary to permanent and is moved from the Hits category into the appropriate feature category. The feature can now be saved to the local GenomBench database as part of the fragment file.

**Showing Sim4 and Spidey Results in the Alignment Tab**

Sim4 and Spidey alignments can be viewed in the Alignment tab of the Information Pane. You can show such alignments by dragging the analysis hits into the Alignment tab, in the same way other alignments are performed. For more information on performing alignments in the Alignment tab, see page 523.

**Querying Public Databases**

Vector NTI’s Query Application can perform a number of query searches on different public servers directly from GenomBench, including:

- UCSC (Genome Browser)
- UCSC (BLAT Search)
- Ensembl (Genome Browser)
- Ensembl (SSAHA Search)
- PubMed Entrez Search

The Query Application provides a convenient user interface where you can enter your search criteria. It sends your search criteria to the selected public web site where the search is performed. Once the search is completed at the web site, the Query Application returns the results and displays them in the lower portion of the window.
Launching the Query Application

To launch the Query Application, select **Tools > Launch Query App** from the GenomBench menu (Figure 33.45). You can also launch the Query Application directly from the start menu (**Start > Programs > Invitrogen > Vector NTI Suite 10 > Query Application**).

**Note:** If you open the Query Application from GenomBench, it closes when you close GenomBench. If you open the Query Application from the Start menu, it remains open when you close GenomBench.

The Query Application is divided into three main parts: the Navigation Pane, the Query Form, and the Result Pane. A History Pane can also be displayed.

The list of public servers that you can search displays in the Servers List in the Navigation Pane. The Query Form initially displays the list of public servers as well. Expanding one of the folders and selecting a server from the list in the Navigation Pane switches the Query Form to the parameters that are applicable to that server (Figure 33.46). Certain fields in the Query Form are pre-configured and others can be changed to select a condition or value. The Query Form for each server is unique and each is described in Configuring Query Application Parameters on the next page.

The Results, Navigation, and History Panes can be hidden from view by selecting **View > Panes** or by clicking the x in the upper right-hand corner of each. To redisplay, select **View > Panes** again. The Query Form cannot be hidden.
Configuring Query Application Parameters

Each of the available searches has a set of pre-configured parameters that display in the Query Form when you select a search server.

The Query Form can display parameters in a Simple or Advanced form by selecting from the Query Form drop-down menu. The Simple Query Form displays a description of the search and a list of the parameters and values. The Advanced Query Form displays a tabular view of the parameters. The Advanced form also allows you to add and remove additional conditions if permitted by the search parameters. For example, you can add keywords to PubMed and Ensembl Genome Browser searches but you cannot add conditions to Ensembl SSAHA or UCSC searches. The Advance Query Form is shown in the following examples.

UCSC Genome Browser Parameters

The UCSC Genome Browser query interface allows you to query by UCSC freeze dates for different genomes and by keyword. The query interface for the UCSC Genome Browser is shown in Figure 33.47.

Select an organism and freeze data from the Value field in the first row and enter a keyword value in the Value field in the second row. Click the Submit Query button in the Query Form or the button on the toolbar to start the search. Click the button on the toolbar to stop the search. Results are displayed in the Results Pane.

For more information about the UCSC Genome Browser, see the UCSC Genome Bioinformatics web site at [http://genome.cse.ucsc.edu/cgi-bin/hgGateway](http://genome.cse.ucsc.edu/cgi-bin/hgGateway) or click the UCSC Genome Browser hyperlink on the top of the Query Form.

UCSC BLAT Search Parameters

BLAT on DNA is designed to quickly find 40+ base pair sequences of 95% and greater similarity. It may miss more divergent or shorter sequence alignments. It typically finds perfect sequence matches of 33 bases, and can sometimes find matches as low as 22 bases. DNA BLAT works well on primates.
The UCSC BLAT query interface allows you to query by UCSC freeze dates for different genomes, by query type, and by sequence. The query interface for UCSC BLAT is shown in Figure 33.48.

Select the organism and freeze date from the drop-down list in the first row. Select the query type from the drop-down list in the second row. Select how you would like the output sorted from the drop-down list in the third row. Enter the query sequence by either entering it or pasting it into the textbox in the fourth row in FASTA format. Click the Submit Query button in the Query Form or the button on the toolbar to start the search. Click the button on the toolbar to stop the search. Results are displayed in the Results Pane.

For more information on the UCSC BLAT search, see the UCSC Genome Bioinformatics website at: http://genome.cse.ucsc.edu/cgi-bin/hgBlat?command=start&org=Human&db=hg15&hgsid=23397897 or click the UCSC BLAT Search hyperlink on the top of the Query Form.

Ensembl Genome Browser Parameters

The Ensembl genome browser interface allows you to search the Ensembl databases by organism, search field, and keyword. The query interface for the Ensembl Genome Browser is shown in Figure 33.49.

Select the organism database you want to search from the Value field drop-down list in the first row. Select the field you want to search from the Value field drop-down list in the second row. Enter the keyword you want to search in the Value field in the third row. You can also add one or more additional keywords by clicking the Add Condition button. Click the Remove Condition button to remove the bottom row. Click the Submit Query button in the Query Form or the button on the toolbar to start the search. Click the button on the toolbar to stop the search. Results are displayed in the Results Pane.
For more information on the Ensembl Genome Browser, see the Ensembl Genome Browser web site at: http://www.ensembl.org/ or click the Ensembl Genome Browser hyperlink on the top of the Query Form.

Ensembl SSAHA Search Parameters

SSAHA (Sequence Search and Alignment by Hashing Algorithm) is a software tool for fast matching and alignment of DNA sequences. It achieves its fast search speed by converting sequence information into a hash table data structure, which can then be searched rapidly for matches.

The Ensembl SSAHA query interface allows you to search the Ensembl databases by a number of fields. The query interface for Ensembl SSAHA is shown in Figure 33.50.

For each field, except the Sequence field, select a value from the drop-down list in the Value field. For the Sequence field, either enter or paste the query sequence into the textbox in FASTA or plain text format. Click the Submit Query button in the Query Form or the button on the toolbar to start the search. Click the button on the toolbar to stop the search. Results are displayed in the Results Pane.

For more information on SSAHA, see http://www.sanger.ac.uk/Software/analysis/SSAHA/. For more information on the Ensemble SSAHA search, see http://www.ensembl.org/Homo_sapiens/ssahaview or click the Ensembl SSAHA Search hyperlink on the top of the Query Form.

PubMed Entrez Search Parameters

The PubMed Entrez query interface allows you to search for citations by database and by keyword. The query interface is shown in Figure 33.51.
Select the database to search from the drop-down list in the Value field in the first row. Enter a keyword to search for in the Value field in the second row. You can also add one or more additional key words by clicking the Add Condition button. Click the Remove Condition button to remove the bottom row. Click the Submit Query button in the Query Form or the button on the toolbar to start the search. Click the button on the toolbar to stop the search. Results are displayed in the Results Pane.

For more information on PubMed Entrez, see http://www.ncbi.nlm.nih.gov/entrez/query.fcgi or click the PubMed Entrez Search hyperlink on the top of the Query Form.

**Query Application Properties**

Several properties can be customized in the Query Application including URLs to search servers. Selecting Edit > Properties opens the Query Application Properties dialog box, which has two tabs: General and Advanced.

In the General tab:

- In the Common box, enter the number of search results per page in the field. If you change this number after a query is already run, the number of search results per page does not update until the next query is run.
- In the User Interface box, check the boxes to Enable dockable windows, Remember last used server, and Remember window/position sizes.

In the Advanced tab:

- In the Servers Configuration box, check the box to use a custom configuration file. This allows you to configure URLs to servers other than the public servers listed. For more information, contact Invitrogen Customer Support.
- In the Warnings box, check the box to turn off the warning that tells you if you start a new request before an existing query is finished running, the existing query will be stopped.

**Query Application Results**

When a search completes, the Results Pane of the Query Application displays the query results (Figure 33.52):

![Figure 33.52 Query Application results display in the Results Pane](image)

Query result information is organized into columns in the Results Pane. The types of information/columns in the display depend on the query performed. Hyperlinks to hits appear in the
Name column. Click on a link to open the hit in the corresponding Vector NTI Advance application (for example, genomic sequence hits open in GenomBench, PubMed Entrez hits open in Citation Viewer, etc.).

All queries performed in a single session of the Query Application are recorded in the History Pane and are organized under the search that was used to perform them (Figure 33.53). To show the History Pane, select View > Pane > History. Results in the Results Pane are replaced by the results from a subsequent search. To rerun a query, click on the hyperlink in the History Pane to redisplay the results in the Results Pane.

![Figure 33.53 Two queries displayed in the History Pane](image)

Once you have results for at least one query, the following buttons on the toolbar and commands on the Results Pane shortcut menu are available:

- **Open Link/Save to DB** – Opens the selected hit in the corresponding Vector NTI Advance application.

- **Create Shortcut on Desktop** – Creates a link to the selected hit as a shortcut on your desktop.

- **Select All** – Selects all the hits in the Results Pane.

- **Save Selected Results** – Saves the selected hits to a tab-delimited text file.

- **Save All Results** – Saves all hits to a tab-delimited text file.

- **Copy/Copy Link** – Copies the information in each cell for the selected hits including the underlying hyperlink. You can then paste the information into any word processing program.

- **List Searches** – Lists the searches performed in the current session of the Query Application. Click the drop-down list next to either button to advance or go back to results from a different query.

- **Previous/Next Page** – Displays the previous or next page of results if there are more results than will fit on one page.

- **Query Form Pane** – Displays the Query Form Pane.
Sending Fragments to Vector NTI

You can send a displayed fragment, current selection, marked region(s) or selected feature from GenomBench to Vector NTI. To send an item to Vector NTI, make the appropriate mark(s) or selection(s) and select **Tools > Send Sequence to VNTI**. The Launch VNTI Options dialog box displays (Figure 33.54):

![Launch VNTI Options dialog box](image)

**Figure 33.54 Launch VNTI Options dialog box**

In the Launch VNTI Options dialog box, configure the parameters for exporting the item to Vector NTI:

- **Range** – Select the region you want to export to Vector NTI.
  - Current fragment – Entire fragment currently open in GenomBench
  - Selection – Selected area of the current fragment.
  - Mark(s) – Marked area(s) of the current fragment
  - Selected Feature – Selected feature on the current fragment. Feature is spliced according to its sub-regions.
- **Features** – Choose which features you would like to import (no features, all features or selected features only). Partially overlapped features are truncated upon transfer.
- **Description** – Add or modify the description for the exported item.
- **Translate to Protein** – Check the **Translate to Protein** box if you want to translate the DNA sequence to protein upon export.
Click OK to send to Vector NTI. The sent item opens in a Molecule Display window in Vector NTI (Figure 33.55).

Features display for the new molecule according to the configuration you set in the Launch VNTI Options dialog box. From the VNTI Molecule Viewing window, you can save the new molecule to the Vector NTI Database and work with it in Vector NTI as you would any other molecule.

**Note:** Many of the genomic feature type names do not coincide with conventional GenBank feature type names. In this case, the genomic features come into Vector NTI as Miscellaneous features. However, the original feature type is captured in the feature description.

### Saving Sequences

Current fragments or portions of the current fragment can be saved to the local GenomBench database. Saving via the Save/Save As command saves the current fragment, whereas saving via the **Create New Sequence** command allows you to save the current fragment or a portion of it.

### Saving Fragments

To save the entire current fragment, select File > Save or Save As from the menu. The single-tabbed Save to Local Storage dialog box displays. Initially, when you open the Save to Local Storage dialog box, the window is empty. Before you can save to the Local database, you must first create at least one subset in which to store your sequences.

To create a new subset, click the Create New Subset button ( ). A new entry appears in the Subset Pane where you can enter the name of the new subset.

To save the fragment, select a subset and enter the name for the fragment in the Molecule Name box. Click OK to save the fragment.

### Creating New Sequences

To create a new sequence from the current fragment and save it to the local GenomBench database, select File > Create New Sequence or click the Create New Sequence button ( ) on the toolbar. The Create New Sequence dialog box displays with save options at the bottom (Fig-
In the Local Storage tab, create a subset, if necessary (see Saving Fragments above), select the subset and define the sequence name.

In the **Range** and **Features** boxes at the bottom of the dialog box, define the portion and features of the fragment you want to save.

- **Range** – Select the region you want to save (current fragment, selection, mark(s), or selected feature).
- **Features** – Select which features you want to save (no features, all features or selected features only). Partially overlapped features are truncated upon saving.
- **Description** – Enter a description for the saved item.

Click **OK** to save the sequence.

**Note:** When you save a sequence using the Create New Sequence option, the sequence is given a start position. If you create the new sequence using the **Selection** or **Selected Feature** option, the start position is based on the position coordinate on the genomic sequence backbone. If you create the new sequence using the **Marks** option, the start position is set to 1. If the sequence is opened from the Local database, the backbone coordinate is preserved only when saving the whole fragment; in all other cases the start position is set to 1.
Importing and Exporting Sequences

GenomBench provides importing and exporting capability for sequences. Importing is performed using the GenomBench import utility; exporting is done via a menu command in the GenomBench window.

Importing Files into GenomBench

Sequences are imported into the local GenomBench database from either GenBank (.gb) or Sequence (.seq) files using the GenomBench import utility. To launch the GenomBench import utility, select start > programs > invitrogen > vector NTI suite 10 > genombench utilities > genombench import. The WinImport dialog box opens (Figure 33.57):

![GenomBench import utility](image)

If you had the previous version of GenomBench (version 1) installed on your system with Vector NTI Advance (version 8) and you have not already migrated the database, the message in Figure 33.58 displays asking you to migrate your current database so that it is compatible with the new version of GenomBench (version 2). If you plan to use the existing version(s) of a GenomBench database, click cancel in the message, and make a backup copy of the database. To locate all copies of existing databases, select start > programs > invitrogen > vector NTI suite 10 > genombench utilities > ls configuration. Paths to any configured databases are listed in the drop-down menu in the Database Name field. To backup the database, simply copy the .mdb file to a new location.

![Message to migrate the database](image)

To continue with the migration, click OK in the message. The migration proceeds and opens the GenomBench window when finished. Any existing data remains in the database and you can now import sequences to local storage.

In the WinImport dialog box, select the file to import by clicking the Browse button. Clicking the Open button enters the import file name and path into the File to Import box in the WinImport dialog box. Select the subset you want to import the file into from the Subset drop-down list and
click the Start button to begin the import process. The import status is reported in the textbox in the lower half of the WinImport dialog box (Figure 33.59).

The **Update Mode** option updates existing entries in the local GenomBench database with information from imported entries, when an existing entry and imported entry match. GenomBench searches for a match as follows:

- The accession number of the sequence that is being imported is compared to the entry in the local database.
- If the accession numbers match, the sequences are compared.
- If the sequences match, GenomBench updates information, such as description, name, qualifiers, etc., for the existing sequence.
- If the sequences do not match, GenomBench adds the imported sequence as a new sequence in the local database.
- Imported sequences that do not have accession numbers are not compared; they are always added to the local database as new sequences.

To invoke Update Mode, check the **Update Mode** checkbox.
Exporting Sequences to GenBank Files

A current fragment or a portion of the current fragment can be exported from GenomBench to a GenBank (.gb) file. To export the current fragment (portion), select File > Export to GenBank File. The Export Options dialog opens (Figure 33.60).

Figure 33.60 Export Options dialog box

Similar to the Save Options dialog box, in the Export Options dialog box you can define the portion and features of the fragment you want to export:

- **Range** – Select the region you want to export (current fragment, selection, mark(s) or selected feature).
- **Features** – Choose which features you would like to export (no features, all features or selected features only). Partially overlapped features are truncated upon exporting.
- **Description** – Enter a description for the exported item.
- **Translate to Protein** – Check the Translate to Protein option if you want GenomBench to export the translation of the sequence.

When you have set the export options appropriately, click OK. The Export to GenBank File dialog box opens. Navigate to the location where you want to store the exported file, enter a name for the file, and click the Save button to complete the export process.

Printing in GenomBench

Graphics and text can be printed in GenomBench by pane. In the Overview and Feature Panes, the graphics are printed. In the tabs of the Information Pane, the text-based information is printed. To print a pane, select a pane or tab to make it active and then select File > Print or
click the **Print** button ( ) on the toolbar. In the Print dialog box that opens, you can also select which pane you want to print in the **Select Pane** box (Figure 33.61).

![GenomBench print dialog box](image)

*Figure 33.61 GenomBench print dialog box*

Depending on which option is selected in the Select Pane box, different options display in the Print Range and Layout Options boxes. For the Overview and Feature Map panes, you can print all of the graphics or just the visible area. For the Sequence tab, you can print all of the sequence or a specified range only. For the Feature List and Text tabs, the only option is to print all of the information in them.

To preview before printing, click the **Preview** button in the Print dialog box or the **Print Preview** button ( ) on the toolbar.

To print the pane or tab, click **Send to Printer** in the Print dialog box.

**Copying in GenomBench**

Copying in GenomBench can be done via two different methods, Camera and Copy. Both features place data on the clipboard in similar formats. In addition, the Camera feature allows you to place a copied selection in a file.

**Copying by Camera**

The Camera tool copies the selected pane as it is represented, However, you can select options in the Camera dialog box that modify how the copied material is represented.
To copy data, select the pane or tab you want to copy to make it active and select **Edit > Camera** or click the **Camera** button ( ) on the toolbar. In the Camera dialog box (Figure 33.62), you can also select any of the panes or tabs from the **Select Pane** drop-down list.

![Camera dialog used for copying data](image)

In the Camera dialog box, select from the following parameters:

- **Format**: Select between the items being copied, either graphics or text format.
- **Range**: Specify the range of data to be copied. The selection option availability is based on the choice you made in the Select Pane drop-down list.
- **Copy/Save to File**: These options finalize the copy operation.
  - The **Copy** option copies the entire pane contents to the clipboard from which you can paste into any word processing program.
  - The **Save to File** option transfers the camera-copied material to file in the specified format, an .emf or a .jpg file, and representation. In the Save dialog box that opens, specify the destination for the file you are saving.

**Copying by the Copy Command**

The Copy command, invoked by selecting **Edit > Copy**, does not open a dialog box. It simply copies to the clipboard all or a selected portion (such as a feature or sequence fragment) of an activated pane (indicated by the dark blue title bar to the left of the pane).

The following can be copied from the panes listed:

<table>
<thead>
<tr>
<th>Feature Map Pane</th>
<th>Names of selected features copy as tab-delimited text.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Features copy in the internal binary format</td>
</tr>
<tr>
<td>Alignment Tab</td>
<td>A selected range of the alignment copied as text</td>
</tr>
<tr>
<td></td>
<td>Features copy in the internal binary format</td>
</tr>
<tr>
<td>Sequence Tab</td>
<td>Selected range copies from the direct strand in plain text format</td>
</tr>
<tr>
<td>Feature List Tab/Feature Search Results Tab</td>
<td>Table content copies as tab-delimited text</td>
</tr>
<tr>
<td></td>
<td>Features copy in the internal binary format</td>
</tr>
</tbody>
</table>
Introduction

The Vector NTI Analysis Monitor provides a convenient interface for performing the advanced DNA/RNA and Protein molecule analyses included in BioAnnotator and GenomBench, as well as an alternative interface for performing BLAST searches from Vector NTI.

BioAnnotator includes the following advanced analyses:
- PFAM
- ProSite
- BLOCKS
- Proteolytic Cleavage.

GenomBench includes the following advanced analyses:
- Spidey
- Sim4

Note: Analyses that are included with GenomBench or BioAnnotator licenses also become available from the Vector NTI window when these products are licensed. Analyses that are included with GenomBench become available from BioAnnotator when both products are licensed.

Once an analysis is submitted in the Analysis Monitor, it continues to run unless it is aborted, even if the Analysis Monitor application is closed. If Analysis Monitor is closed while an analysis is running, it can be reopened at any time to check the status of the analysis.

Vector NTI functions as a viewer for analysis results retrieved by the Analysis Monitor. The Analysis Results data type enables you to store results from the Analysis Monitor application.

PFAM, ProSite and BLOCKS Databases

PFAM, ProSite and BLOCKS databases used for the above analyses are provided in the Vector NTI Advance 10 installation. They are located in the AM subdirectory of the VNTI Database directory. For example, if you perform a default Vector NTI Advance 10 installation, these data-
base files are located in at C:\vnti database\am. The PFAM, ProSite and BLOCKS database files are named pfam, prosite.dat and blockz.ddb, respectively.

The PFAM database file is from the PFAM HMM library (Pfam_ls or Pfam_fs) available at http://pfam.wustl.edu/.

The prosite.dat file is from ftp://us.expasy.org/databases/prosite/release/.

In the case of the BLOCKS database, the blocks.dat file from ftp://ftp.ncbi.nih.gov/repository/blocks/unix/ is converted to binary code to create the blockz.ddb database file that is used for the BLOCKS analysis feature. Your Vector NTI Advance 10 installation includes a conversion program called blockZ.exe that reads in a blocks.dat text file and outputs a blockz.ddb binary file that can be used for the BLOCKS analysis.

You can update your PFAM, ProSite or BLOCKS database file by downloading the appropriate file from the corresponding above-mentioned location, converting the file (in the case of blocks.dat) and replacing the existing database file in the VNTI Database AM subdirectory with the downloaded/converted file.

Launching Advanced Analyses

You can launch an advanced analysis for the active molecule in the BioAnnotator or Vector NTI window (see above note) by choosing Analyses > <BioAnnotator or GenomBench Tools> <advanced analysis tool name> from the drop-down menu in the application window. In the GenomBench window, choose Tools > Align cDNA with Genomic DNA using <analysis tool>. When the Analysis Monitor is launched in this way, its interface is specific for the analysis launched.

Choosing Analyses > Run Analysis Monitor from the main Vector NTI workspace is a second method for launching Analysis Monitor for the active/selected molecule. In this case, a preliminary dialog box displays showing the active/selected molecule information and the advanced analysis options for that molecule type. Select the analysis type you want to run, specify a name for the analysis and click the Apply button to launch Analysis Monitor for the indicated molecule and analysis. The Run Analysis Monitor command is available from the Analyses menus in BioAnnotator and in the Vector NTI Molecule Viewing and Local Database windows.

You can also launch the Analysis Monitor from the Start menu by choosing Start > Programs > Invitrogen > Vector NTI Suite 10 > Analysis Monitor (assuming a default VNTI installation was performed). In this case, the Analysis Monitor dialog box opens with an empty Parameters Pane. Launching Analysis monitor is this way is useful for performing advanced analyses on molecules saved to an external location in GenBank/GenPept format, for starting previously defined analyses that have not yet been run and for viewing the status of currently running analyses (see the next section).

To load in a molecule file and analysis type, choose Analysis > <molecule type (DNA/RNA or protein)> <advanced analysis option> from the Analysis Monitor menu and navigate to the molecule file to be analyzed in the Open dialog box. The appropriate interface opens in the Analysis Monitor for the selected molecule file. To load in a previously saved analysis, choose Analysis > Load from File from the menu.

Configuring parameters for the different analyses is discussed in the sections below.
Elements of the Analysis Monitor Workspace

The Analysis Monitor interface consists of Title, Menu and Tool bars, a Parameters Pane for configuring analysis parameters and an Analysis List Pane that lists the analyses performed and their stage of completion (Figure 34.1):

![Figure 34.1 Analysis Monitor interface](image)

The Analysis Monitor title bar reflects the selected analysis and name of the molecule being analyzed. The menu/toolbar contains commands for loading analyses and molecules from files, saving analyses, basic copying and pasting and for viewing information about the Analysis Monitor program.

In the Parameters Pane of the Analysis Monitor interface, you can configure analysis-specific parameters as described in the next sections. For blue-colored parameter labels, you can hover the cursor over the label and click when the hand icon appears to open a tip dialog box that describes the parameters in that section (Figure 34.2):

![Figure 34.2 Tool tip for Spidey Parameters section](image)
The Analysis List Pane of the Analysis Monitor shows the list of analyses that have been configured/ performed. The icon at the left of each listed analysis indicates the stage of completion of the analysis as follows:

<table>
<thead>
<tr>
<th>Icon</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Analysis added, but not yet initiated</td>
</tr>
<tr>
<td></td>
<td>Analysis in the process of loading</td>
</tr>
<tr>
<td></td>
<td>Analysis running</td>
</tr>
<tr>
<td></td>
<td>Analysis completed successfully</td>
</tr>
<tr>
<td></td>
<td>Analysis aborted</td>
</tr>
</tbody>
</table>

Table 34.1 Analysis stage of completion icons

For analyses that have been added, but not yet initiated, double-clicking on the analysis name in the list loads the analysis into the upper portion of the dialog box, so that you can configure and run the analysis.

To initiate the analysis currently displayed in the Parameters Pane, click the Submit button.

**Configuring Advanced Analysis Parameters**

In the Analysis Monitor interface, you can configure parameters to customize the advanced analysis option you have selected. The different types of searches/analyses and their parameters are discussed in this section.

**Sim4**

Sim4 is a similarity search tool for aligning an expressed DNA transcript (i.e. an EST, cDNA or mRNA) with a template (genomic sequence). Sim4 employs the following multi-stage BLAST-based technique:

1. Sim4 detects all possible exact matches of W-mers between the two sequences and extends them to maximal scoring gap-free segments (exon-cores).
2. Exon-cores are extended into the adjacent, as-yet-unmatched fragments using greedy alignment algorithms. Heuristics are used to favor configurations that conform to the splice-site recognition signals. If necessary, the process is repeated with less stringent parameters on the unmatched fragments.

Sim4 functions similarly to BLAST, but performs a more thorough mRNA alignment search.
The Analysis Monitor Parameters Pane for the Sim4 analysis tool appears as follows (Figure 34.3):

**Figure 34.3 Sim4 parameters**

**Sequence to Analyze**
- **Name** – Specifies the name of the molecule sequence you want to analyze.

**Sim4 Parameters**
- **Word Size** – The number of DNA bases in a size unit. The larger the word size, the faster and less sensitive Sim4 becomes.
- **Limit of Score Drop-off** – The trigger for stopping ungapped extension.
- **Search for Small Exons** – If **Yes** is selected, an additional search for small exons is performed.
- **Diagonal Distance** – The upper boundary of diagonal distance within consecutive HSPs in an exon.
- **Allow Ambiguity Characters** – If checked, allows the following ambiguity characters: ABCDGHKMNRSTVWXYZ. If unchecked, only the following characters are allowed: ACGTNX.
- **Weight Factor for Linking HSPs** – The multiplication factor used when calculating the score of a chain of HSPs.
- **Use Defaults** – Resets Sim4 parameters to the original default values.

**HSP Score Threshold**
- **First Stage** – The threshold for HSPs for the first stage of comparison. If no value is specified, the default value is used.
- **Second Stage** – The threshold for HSPs when aligning the extended-and-unmatched portions. If no value is specified, the default value is used.

**Against Sequences**
Select the subset of genomic sequences, mRNAs or ESTs to search against.

**Strands**
Select the strand(s) to search. The **Both** option searches both strands and reports the best result.
Spidey

Spidey is an mRNA-to-genomic alignment program. It takes as input a single genomic sequence and a set of mRNA accessions or FASTA sequences. All processing is done one mRNA sequence at a time. The first step for each mRNA sequence is a high-stringency BLAST against the genomic sequence. The BLAST alignments are sorted by score and then assigned into windows by a recursive function.

After the genomic windows are constructed, the initial BLAST alignments are freed and another BLAST search is performed, this time with the entire mRNA against the genomic region defined by the window, and at a lower stringency than the initial search. Spidey then uses a greedy algorithm to generate a high-scoring, non-overlapping subset of the alignments from the second BLAST search. This consistent set is analyzed carefully to make sure that the entire mRNA sequence is covered by the alignments.

Once the mRNA is completely covered by the set of alignments, the boundaries of the alignments are adjusted so that the alignments abut each other precisely and so that they are adjacent to good splice donor and acceptor sites. To position the exon boundaries, the adjacent exon alignment overlap region plus a few base pairs on each side is examined for splice donor sites, using functions that have different splice matrices depending on the organism chosen. The top few splice donor sites (by score) are then evaluated as to how much they affect the original alignment boundaries. The site that affects the boundaries the least is chosen, and is evaluated as to the presence of an acceptor site. The alignments are truncated or extended as necessary so that they terminate at the splice donor site and so that they do not overlap.


The Analysis Monitor Parameters Pane for the Spidey analysis tool appears as follows (Figure 34.4):

![Figure 34.4 Spidey parameters](image)

**Genomic Sequence to Analyze**

- **Name** – Specifies the name of the molecule’s genomic sequence you want to analyze.
- **Organism** – Select the organism type for your genomic sequence.

**Spidey Parameters**

- **Minimum mRNA Genomic Identity** – Percent identity cutoff for gene models.
- **Minimum Length of mRNA Covered** – mRNA length coverage cutoff in percent for gene models.
- **Limit Number of Gene Models to** – Sets the maximum number of gene models Spidey analysis returns.
● **E-value**
  ○ **1st Pass** – The E-value cutoff for the initial high-stringency alignment. The higher the value, the less stringent and faster the run.
  ○ **2nd Pass** – The E-value cutoff for the low-stringency BLAST search within a genomic window, based on the high-stringency result.
  ○ **3rd Pass** – The E-value cutoff for a very low-stringency BLAST search to find hits for mRNA gaps.

● **Divergent Sequences** – If checked, search parameters are adjusted to tolerate mismatches and gaps for inter-species alignment.

● **Use Large Intron Sizes** – If checked, much larger maximal intron sizes are allowed. Checking this option increases computation time significantly.

● **Use Defaults** – Sets Spidey parameters to original default values.

**Analyze Against mRNA Sequences**

Select the Vector NTI Database subset of mRNA molecules you want to use for the analysis.

**BLOCKS**

Blocks are multiply aligned ungapped segments corresponding to the most highly conserved regions of proteins. The BLOCKS database is built by taking lists of functionally related proteins and using automated methods to locate common sequence motifs. It is comprised of blocks in representative protein families.

The Analysis Monitor Parameters Pane for the BLOCKS analysis tool appears as follows (Figure 34.5):

![Figure 34.5 BLOCKS parameters](image)

**Protein to Analyze**

● **Range Only** – Designate the part of the molecule’s sequence you want to analyze. If unchecked, the entire sequence is analyzed.

**PFAM**

PFAM is a database of multiple alignments of protein domains or conserved protein regions. The alignments represent evolutionary conserved structure, which can have implications for a protein’s function. Profile hidden Markov models (profile HMMs) built from the PFAM alignments can be very useful for automatically recognizing that a new protein belongs to an existing protein family, even if the homology is weak. Unlike standard pairwise alignment methods (e.g. BLAST, FASTA), PFAM HMMs deal sensibly with multi-domain proteins. The PFAM tool in GenomBench takes a protein sequence as input and compares it against a database of HMMs, producing a set of alignments with the best scoring HMMs.

For more information on PFAM and HMMs, see [http://pfam.wustl.edu/help.shtml](http://pfam.wustl.edu/help.shtml).
The Analysis Monitor Parameters Pane for the PFAM analysis tool appears as follows (Figure 34.6):

**Figure 34.6 PFAM parameters**

**PFAM Settings**

- **Algorithm** – Select the algorithm to use for per-sequence scores.
- **Turn Off Postprocessing** – If checked, post-processing is not performed. This usually results in higher sensitivity, but lower specificity.
- **Use XNU Filtering** – If checked, the target protein sequence is checked for tandem repeats.
- **Limits**
  - **Scoring Limit** – Check this option and specify a value to limit the number of best scoring domains in the alignment output.
  - **E-value Cutoff** – Hits with E-values less than the set threshold are shown. The default value is 10.
  - **Bit Score Cutoff** – Check this option and specify a value to show only those hits with bit scores greater than this threshold. If unchecked, the threshold is controlled by E-value rather than bit score.

**Proteolytic Cleavage**

The Proteolytic Cleavage tool reports digestion fragments for a molecule with the designated cleavage reagent. You can use one of the several common preset cleavage reagents or use a reagent that you custom define.

The Analysis Monitor Parameters Pane for the Proteolytic Cleavage analysis tool appears as follows (Figure 34.7):

**Figure 34.7 Proteolytic Cleavage parameters**
Protein Molecule
- **Range Only** – Designate the part of the molecule’s sequence you want to analyze. If unchecked, the entire sequence is analyzed.

Results
- **Best <#> Fragments** – Select this option to display the designated number of most likely fragments
- **Only Above This Probability** – Select this option to display only the most likely fragments above the threshold you designate.

Results Sorted By
- **Probability** – Sort fragments by probability in the Text Pane of the Molecule Viewing window.
- **Fragment Weight** – Sort fragments by weight in the Text Pane of the Molecule Viewing window.
- **Position on the Molecule** – Order fragments according to their position on the molecule, from the start of the molecule to the end.

Cleavage Reagents
- **Select a Cleavage Reagent** – Highlight the cleavage reagent you want to use for the analysis.
- **New** – Click the **New** button to launch the Edit Reagent dialog box (Figure 34.8) where you can define a new cleavage reagent. When you open the Edit dialog box via the New button, the dialog box is empty, allowing you to define a new reagent from scratch. If you want to define a new reagent by modifying an existing reagent, highlight the reagent in the list and click the **Edit** button (see below).
- **Edit** – Highlight the reagent you want to modify and click the **Edit** button. The Edit Reagent dialog box (Figure 34.8) opens with the highlighted reagent’s cleavage pattern parameters, allowing you to edit them.

![Figure 34.8 Edit Reagent dialog box](image)

- Add additional positions (if any) and recognition or exception patterns according to IUPAC convention (see Appendix C) for the appropriate positions in the upper sections of the dialog box.
- Use curly braces {} to group disallowed amino acids and brackets [ ] to group allowed amino acids for a position.
- Click the **Add** button to add the new recognition or exception pattern to the list in the lower portion of the dialog box.
To delete a recognition or exception pattern, highlight it in the list and click the **Delete** key in that section.

- Name (in the case of a new reagent) the new configuration and click the **OK** button to save the new or modified reagent to the reagent list.

- **Remove** – Highlight a reagent on the list and click **Remove** button to delete it from the list.

- **Use Defaults** – Restores parameter values to Analysis Monitor defaults.

### ProSite

ProSite is a database of biologically significant protein patterns and profiles that help to reliably identify the function of uncharacterized proteins. A ProSite search can determine known protein families and domains to which a newly translated polypeptide sequence may belong.

An unknown amino acid sequence may be too distantly related to proteins of known structure to demonstrate any similarity in an overall sequence alignment. A ProSite search, on the other hand, may reveal in the sequence a cluster of similar residues, called a pattern, motif, signature or fingerprint, that suggest a relationship with other known protein families. These specific domain(s), highly conserved in both structure and sequence, can provide important clues to the function of newly discovered proteins.


The Analysis Monitor Parameters Pane for the ProSite analysis tool appears as follows (Figure 34.9):

![ProSite parameters](image)

**Exclude Frequent Patterns from Search**

Check the boxes of the common patterns you would like to exclude from the search.

**User’s Patterns**

- **Add** – Type the pattern in ProSite format in the drop-down box and click the **Add** button to add the new pattern to the user-defined pattern list.

Patterns are described using the following ProSite conventions:

- The standard IUPAC one-letter codes for the amino acids are used.
- The symbol x is used for a position where any amino acid is accepted.
- Ambiguities for acceptable amino acids for a given position are listed in square brackets [ ]. For example, [ALT] stands for Ala or Leu or Thr.
Ambiguities for the amino acids that are NOT accepted at a given position are indicated by listing them in curly braces \{\}. For example, \{AM\} indicates any amino acid EXCEPT Ala and Met.

- Each element in a pattern is separated from its neighbor by the dash character -.
- Repetition of an element of the pattern can be indicated by following that element with a numerical value or a numerical range between parenthesis. Examples: \(x(3)\) corresponds to \(x-x-x\) and \(x(2,4)\) corresponds to \(x-x\) or \(x-x-x\) or \(x-x-x-x\).
- When a pattern is restricted to the N- or C-terminal of a sequence, the pattern starts with a < symbol or ends with a > symbol, respectively. In some rare cases (e.g. PS00267 or PS00539), > can also occur inside square brackets for the C-terminal element. For example, \(F-[GSTV]-P-R-L-[G>]\) means that either \(F-[GSTV]-P-R-L-G\) or \(F-[GSTV]-P-R-L>\) are considered.

Examples:

- \([AC]-x-V-x(4)-{ED}\)
- This pattern is translated as: [Ala or Cys]-any-Val-any-any-any-any-any but Glu or Asp]
- \(<A-x-[ST](2)-x(0,1)-V\)
- This pattern, which must be in the N-terminal of the sequence (‘<’), is translated as: Ala-any-[Ser or Thr]-[Ser or Thr]-any or none)-Val

- **Remove** – Highlight a user-defined pattern in the list and click the **Remove** button to delete it from the list.
- **Remove All** – Removes all user-defined patterns from the list.
- **Check Format** – Highlight a user-defined pattern in the list and click the **Check Format** button to verify that it is in proper ProSite format. A message displays indicating whether or not the pattern is in proper format.

**Do Not Show Patterns with More Than <#> Sites**

Check the box and designate the maximum number of sites for shown patterns.

**BLAST Search in Analysis Monitor**

BLAST (Basic Local Alignment Search Tool) searches explore, using a DNA or protein query sequence, available sequence databases for sequence similarity. BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. BLAST scores have a well-defined statistical interpretation, making real matches easy to distinguish from random background hits. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity. (Altschul, et al. 1990).

When launching BLAST Searches from Analysis Monitor, the BLAST interface reflects the NCBI workspace and uses preset parameters and defaults providing a uniform approach to sequence analysis. For detailed information on BLAST search types, settings, parameters, search databases, etc., see http://www.ncbi.nlm.nih.gov/Sitemap/index.html#BLAST2.x

To launch BLAST Searches for the active/selected molecule using Analysis Monitor, choose **Analyses > Run Analysis Monitor** from the main Vector NTI workspace. In this case, a preliminary dialog box displays showing the active/selected molecule information and all of the advanced analysis options for that molecule type. Select the BLAST Search type you want to run, specify a name for the analysis and click the **Apply** button to launch Analysis Monitor for the indicated molecule and analysis. For more information on the different types of BLAST Searches, see Database Selection Parameters, page 576.
Operations on Analyses

There are a number of operations you can perform on the analyses listed in the Analysis List Pane. Most operation commands are available on the Analysis drop-down menu and on the shortcut menu launched by right-clicking on the analysis of interest in the Analysis List Pane.

Starting an Analysis

There are several methods for starting an analysis, depending on the completion stage of the analysis.

**Current Analyses**

To initiate the analysis currently displayed in the Parameters Pane, click the **Submit** button.

**Failed Analyses**

For failed analyses, right-click on the analysis in the Analysis List Pane and choose **Start** from the shortcut menu or highlight the failed analysis and choose **Analysis > Start** from the menu to start the analysis.

**NewlySubmitted Analyses**

To start analyses that have been configured, but not yet run (those shown in the Analysis List Pane with the icon), you must first open the analysis in the Parameters Pane. There are several ways to open such an analysis in the Parameters Pane:

- Double-click on the analysis listing in the Analysis List Pane
- Right-click on the analysis listing and choose Edit or
- Highlight the analysis listing and choose **Analysis > Edit** from the menu.

Once you have opened the analysis in the Parameters Pane, reconfigure the parameters, if necessary, and click the **Submit** button.

Stopping an Analysis

You can stop an analysis that is running by highlighting the analysis in the Analysis List Pane and choosing **Analysis > Stop** from the menu or by right-clicking on the analysis and choosing **Stop** from the shortcut menu.

**Editing, Copying and Deleting**

To open a newly submitted analysis for editing, right-click on the analysis in the Analysis List Pane and choose **Edit** from the shortcut menu or highlight the analysis and choose **Analysis > Edit** from the drop-down menu. The analysis is opened in the Parameters Pane so you can edit its parameters and start the analysis.

**Note:** In the case of completed analyses, you can open an analysis in the Parameters Pane for viewing by highlighting the analysis and choosing **Analysis > Read** from the menu or by choosing the same command from the shortcut menu.

You can make a copy of an analysis by highlighting the analysis and choosing **Make Copy** from the shortcut or Analysis menus. The copy is entered in the Analysis List Pane with the default name of *Copy of <copied analysis name>*. You can perform the same operations on the copy that you can on other analyses.

To delete an analysis highlight it in the Analysis List Pane and choose **Delete** from the shortcut or Analysis menus.
Saving

For completed analyses, you have two saving options, both accessible for the highlighted analysis from either the Analysis drop-down or shortcut menus:

- **Save to File** – Allows you to save the analysis to the hard drive or other location in either an Analysis Archive file or General Format file, both of which can be opened in Vector NTI.

- **Save to DB as Analysis Result** – Allows you to save the analysis to the Analysis Results section of the Vector NTI Database.

Sending to Other Vector NTI Advance Applications

For completed analyses, you can send Protein or DNA/RNA molecule analyses from Analysis Monitor directly to Vector NTI and DNA/RNA analyses directly to GenomBench. When an analysis is sent to Vector NTI or GenomBench, it opens with the molecule in the viewing window (see the next section).

**Sending Analyses to Vector NTI**

To send an analysis directly to Vector NTI, highlight the analysis in the Analysis List Pane and choose **Analysis > Send to Vector NTI** or right-click on the analysis and choose **Send to Vector NTI** from the shortcut menu. The analyzed molecule opens with the sent analysis displayed (see **Viewing Analysis Results in Vector NTI**, page 571).

**Note:** If an analysis was launched from Vector NTI or BioAnnotator, double-clicking the completed analysis in the Analysis Monitor Analysis List Pane sends the analysis to Vector NTI.

**Sending Analyses to GenomBench**

To send an analysis directly to GenomBench, highlight the analysis in the Analysis List Pane and choose **Analysis > Send to GenomBench** or right-click on the analysis and choose **Send to GenomBench** from the shortcut menu. The analyzed sequence opens in GenomBench. (see **Viewing Analysis Results in GenomBench**, page 572).

**Note:** If an analysis was launched from GenomBench, double-clicking the completed analysis in the Analysis Monitor Analysis List Pane sends the analysis to GenomBench.

Viewing Analysis Results

DNA/RNA and Protein analyses can be viewed in Vector NTI in a Molecule Viewing window for the analyzed molecule. DNA/RNA analyses can be viewed in GenomBench.

Viewing Analysis Results in Vector NTI

In Vector NTI, completed analysis results are shown in the Molecule Viewing window of the analyzed molecule. Completed analysis results open with their associated molecule in a Molecule Viewing window when you send Analysis Results to Vector NTI (see above), when you open saved Analysis Results from the Vector NTI Database or when you open a molecule from the Database that has associated Analysis Results.

Details of the completed analysis are listed in the External Analyses folder in the Text Pane and graphical representations of the analysis results appear in the Graphics and Sequence Panes (Figure 34.10).
**Note:** The Analyses section in Display Setup must be turned on and configured appropriately before Analysis Results are shown in the Molecule Viewing window (see page 206).

![Figure 34.10 BLOCKS and PFAM Analysis results shown in Molecule Viewing window](image)

Several operations on analyses/features are available from shortcut menus launched by clicking on the folder for the specific analysis or a hit folder within the specific analysis folder:

- **Convert to Feature(s)** – (hit shortcut menu) Converts the highlighted hit to a feature on the molecule.
- **Convert to Feature(s)** – (analysis shortcut menu) Converts all hits from the analysis to features on the molecule.
- **Hide <analysis type> of <molecule name>** – (analysis shortcut menu) Removes the display for that analysis from the Molecule Viewing window.

**Viewing Analysis Results in GenomBench**

Completed Analysis Results can be shown in GenomBench. When you send results to GenomBench, as described above, they open in the GenomBench workspace with the Sequence tab.
displayed in the Information Pane (Figure 34.11) and hits displayed in the Feature Map Pane and Features tab of the Information Pane.

Figure 34.11 Spidey Analysis Results displayed in GenomBench

For more information about working with Sim4 and Spidey results in GenomBench, see Chapter 33.
Introduction

BLAST (Basic Local Alignment Search Tool) searches explore, using a DNA or protein query sequence, available sequence databases for sequence similarity. BLAST programs have been designed for speed, with a minimal sacrifice of sensitivity to distant sequence relationships. BLAST scores have a well-defined statistical interpretation, making real matches easy to distinguish from random background hits. BLAST uses a heuristic algorithm that seeks local as opposed to global alignments and is therefore able to detect relationships among sequences that share only isolated regions of similarity. (Altschul, et al. 1990).

For detailed information on BLAST search types, settings, parameters, search databases, etc., see http://www.ncbi.nlm.nih.gov/Sitemap/index.html#BLAST2.x

Note: Blast Searches can be also be launched from the main Vector NTI workspace by using the Analyses menu option in the Main Menu. For more information regarding Blast Search in Analysis Monitor, see Chapter 34.

BLAST Search Dialog Box

The BLAST search engine in Vector NTI initiates the search through a Vector NTI dialog box, launching the search through the NCBI (National Center for Biotechnology Information) server at the NCBI website (http://www.ncbi.nlm.nih.gov/blast/).

To open the BLAST Search window, select Start > Programs > Invitrogen > Vector NTI Suite 10 > BLAST Search (assuming a default Vector NTI installation procedure was performed). You
will need to choose the server for the search in the small dialog box that opens (Figure 35.1). (If you’re not sure, select NCBI BLAST Server).

Click OK, opening the following dialog box:

The BLAST Search dialog box (Figure 35.2) contains a small menu bar and toolbar with options relating to search, viewing and help functions. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

This dialog box is divided into three sections: the Database Selection bar, the Sequence and Parameters Pane and the Results Pane.

Database Selection Parameters

**BLAST Search Type** – In the Vector NTI BLAST Search dialog box, there are three search types from which to choose in the drop-down menu:

- **BLAST 2.0** – Basic Local Alignment Search Tool. BLAST 2.0 is a version of BLAST that permits gaps in the alignments it produces.

- **PSI & PHI BLAST**
  - **PSI-BLAST** - Position Specific Iterated BLAST is a program for searching protein databases using protein queries, in order to find other members of the same protein family.
  - PSI-BLAST uses an iterative search in which sequences found in one round of searching are used to build a score matrix for the next round of searching. Highly conserved positions receive high scores and weakly conserved positions receive scores near zero. All statistically significant alignments found by BLAST are combined into a multiple alignment, from which the position-specific score matrix is con-
structured. This matrix is used to perform a second, etc. BLAST search and the results of each "iteration" are used to refine the profile. The process may be iterated until no new alignments are found. This iterative searching strategy results in increased sensitivity.

- **PHI-BLAST** - Pattern Hit Initiated BLAST is a program for searching a protein database using a protein query; it seeks only alignments that preserve a specified pattern contained within the query.

- PHI-BLAST combines matching of regular expression patterns with a Position Specific iterative protein search to locate other protein sequences which both contain the regular expression pattern and are homologous to a query protein sequence.

- **MegaBLAST** – This program uses a "greedy algorithm" (Webb Miller et al., J Comput Biol 2000 Feb-Apr;7(1-2):203-14) for nucleotide sequence alignment searches and concatenates many queries to save time spent scanning the database. It is optimized for aligning sequences that differ slightly and is up to 10 times faster than more common sequence similarity programs. It can be used to swiftly compare two large sets of sequences against each other. MegaBLAST permits searching with batches of ESTs or with large cDNA or genomic sequences.

**Program** – In the drop-down menu, specify the type of database search to be performed. The search options are summarized as follows:

- **blastn** - compares a nucleotide query sequence against a nucleotide sequence database
- **blastp** - compares an amino acid query sequence against a protein sequence database
- **blastx** - compares a nucleotide query sequence translated into all reading frames against a protein sequence database
- **tblastn** - compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands)
- **tblastx** - compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database. *This program cannot be used with the nr database.*

**Database** – In the drop-down menu, select the GenBank database type the query sequence is to be searched against. Many of the potential database type options are described below:

<table>
<thead>
<tr>
<th>Menu Identifier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>nr</td>
<td>Peptide Sequence Database: All non-redundant GenBank sequences and CDS translations. Nucleotide Sequence Database: All GenBank+EML+PDB sequences (no EST, STS, GSS or phase 0, 1 or 2 HTGS sequences). No longer non-redundant.</td>
</tr>
<tr>
<td>EST (+ human, mouse, others)</td>
<td>Nucleotide Sequence Database: EST (Expressed Sequence Tags). Others includes all organisms except mouse and human</td>
</tr>
<tr>
<td>GSS</td>
<td>Nucleotide Sequence Database: Genome Survey Sequence, includes single-pass genomic data, exon-trapped sequences, and Alu PCR sequences.</td>
</tr>
<tr>
<td>HTG</td>
<td>Nucleotide Sequence Database: Unfinished High Throughput Genomic Sequences.</td>
</tr>
<tr>
<td>Patents</td>
<td>Protein sequences from the Patent division of GenBank.</td>
</tr>
</tbody>
</table>

*Table 35.1 GenBank database options*
Sequence/Parameters Pane Tabs

There are four tabs below the database selection section of the screen in the Sequence/Parameters Pane, the **Sequence Tab**, the **Parameters Tab**, the **PSI-PHI Tab** and the **MegaBLAST Tab**.

**Sequence Tab**

On this tab you can enter raw sequence (only) pasted from another application or select Search > Open and locate the file for a sequence you wish to submit. If the sequence is not in raw format, copy it to Notepad, which eliminates all characters except the text itself. Select and copy the sequence from the Notepad and paste it into the text box on the **Sequence** tab. If the BLAST Search module is launched from a Vector NTI Advance application, the active molecule sequence is pasted into the **Sequence** tab.

**Parameters Tab**

On the **Parameters Tab** (Figure 35.3), you can define parameters for the BLAST search.

### Table 35.1 GenBank database options (continued)

<table>
<thead>
<tr>
<th>Menu Identifier</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mito</td>
<td>Nucleotide Sequence Database: Mitochondrial DNA sequences.</td>
</tr>
<tr>
<td>vector</td>
<td>Nucleotide Sequence Database: Vector subset of GenBank.</td>
</tr>
<tr>
<td>E. coli</td>
<td>Peptide Sequence Database: <em>Escherichia coli</em> genomic CDS translations. Nucleotide Sequence Database: <em>Escherichia coli</em> genomic nucleotide sequences.</td>
</tr>
<tr>
<td>month</td>
<td>Peptide Sequence Database: All new or revised GenBank CDS translation + PDB + SwissProt + PIR + PRF released in the last 30 days. Nucleotide Sequence Database: All new or revised GenBank+EMBL+DDBJ+PDB sequences released in the last 30 days.</td>
</tr>
<tr>
<td>Drosophila genome</td>
<td>Peptide Sequence Database: Drosophila genome proteins provided by Celera and Berkeley Drosophila Genome Project. Nucleotide Sequence Database: Drosophila genome provided by Celera and Berkeley Drosophila Genome Project.</td>
</tr>
<tr>
<td>PDB</td>
<td>Peptide and Nucleotide Sequence Database: Sequences derived from the 3-dimensional structure from Brookhaven Protein Data Bank.</td>
</tr>
<tr>
<td>alu</td>
<td>Peptide Sequence Database: Translations of select Alu repeats from REPBASE. Nucleotide Sequence Database: Select Alu repeats from REPBASE.</td>
</tr>
<tr>
<td>dbsts</td>
<td>Nucleotide Sequence Database: Database of GenBank+EMBL+DDBJ sequences from STS Divisions.</td>
</tr>
</tbody>
</table>
Note: The parameters you can set on this tab are almost identical to parameters for the various BLAST searches at the NCBI website: [http://www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html](http://www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html). For more information regarding these parameters than is detailed here, refer to the website.

![Figure 35.3 Parameters tab of the BLAST Search dialog box](image)

**Expect**

The statistical significance threshold for reporting matches against database sequences. The default value of 10 means that in a database of the current size, 10 matches would be expected merely by chance (stochastic model of Karlin and Altschul, 1990.) Hits showing a statistical significance greater than the Expect threshold are not reported. Increasing the E value above 10 produces a larger list with more low-scoring hits (chance matches). Lower expectation value thresholds are more stringent, leading to fewer chance matches being reported.

If your query peptide or nucleotide sequence is short, you might want to increase the Expect value. Because a short query is more likely to occur by chance in the database, even a perfect match can have low statistical significance and may not be reported. Increasing the E value lets you look farther down the hit list and see matches that would normally be discarded because of low statistical significance.

**Word Size**

Word size is roughly the minimal length of an identical match an alignment must contain if it is to be found by the algorithm.

Mega BLAST is most efficient with word sizes 16 and larger, although word size as low as 8 can be used. If the value $W$ of the word size is divisible by 4, it guarantees that all perfect matches of length $W + 3$ will be found and extended by Mega BLAST search, however perfect matches of length as low as $W$ might also be found, although the latter is not guaranteed. Any value of $W$ not divisible by 4 is equivalent to the nearest value divisible by 4 (with $4i + 2$ equivalent to $4i$).

**Filters**

- **Low-complexity** – This filter masks off segments of the query sequence that have low compositional complexity (as determined by the SEG program of Wootton & Federhen, Computational Chemistry, 1993). Regions with low-complexity sequence can create problems in sequence similarity searching by producing artificial hits, sequences that are not truly related. Such hits can produce high scores because of the presence of low-complexity regions.

- **Human Repeats** – This option masks Human repeats and is especially useful for human sequences that may contain these repeats. This option is still experimental and may change in the near future.

- **Mask for Lookup** – This option masks only for purposes of constructing the lookup table used by BLAST. The BLAST extensions are performed without masking. This option is still experimental and may change in the near future.
• **Mask Lower Case** – With this option selected you can cut and paste a FASTA sequence in upper case characters and denote areas you would like filtered with lower case. This allows you to customize what is filtered from the sequence during the comparison to the BLAST databases.

**Note:** For a filter choice of **none**, leave all filter checkboxes unchecked. Check multiple filter checkboxes to combine filters.

**Genetic Codes**

Genetic code to be used in blastx and tblastx translation of the query.

**Other Advanced Options**


**Limit by Entrez Query**

This section lets you set parameters to limit your database search according to organism by selecting from a drop-down list of common organisms in GenBank and/or by entering an organism name (or multiple names in conjunction with the AND or OR operator) in the form, *Genus species*.

**Matrix**

This field permits you to specify the scoring matrix for blastp, blastx, tblastn and tblastx. The default matrix is BLOSUM 62. The valid alternative choices include PAM 30, 70, and BLOSUM 45 and 80. No alternate scoring matrices are available for blastn.

**Composition Based Statistics**

BLAST and PSI-BLAST now permit calculated E-values to take into account the amino acid composition of the individual database sequences involved in reported alignments. This improves E-value accuracy, thereby reducing the number of false positive results. For more information on Composition Based Statistics, go to [http://www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html#composition_based_statistics](http://www.ncbi.nlm.nih.gov/blast/html/blastcgihelp.html#composition_based_statistics).

**Gap Costs**

Gap Costs are the penalty to open a Gap and penalty to extend a Gap. Increasing the Gap Costs decreases the number of Gaps introduced in the alignment.

**PSI-PHI Tab**

The PSI-BLAST search algorithm compares a single protein sequence to a protein database using the gapped BLAST algorithm. First, a multiple local alignment is calculated, followed by a profile alignment, with the original query molecule serving as the template. The profile is then compared to the protein database, again seeking local alignments and producing a profile from the alignment. This iteration process can be repeated an arbitrary number of times.

PHI-BLAST allows you to specify a pattern for the search.
The PSI-PHI Tab (Figure 35.4) allows you to specify additional parameters for PSI-BLAST and PHI-BLAST searches.

**Position Specific Score Matrix**

**Inclusion Threshold**
- The statistical significance threshold for including a sequence in the model used by PSI-BLAST to create the Position Specific Score Matrix on the next iteration.

**PHI Pattern**
- The pattern to be used for the search. The syntax for patterns in PHI-BLAST follows the conventions of PROSITE (valid protein characters for PHI-BLAST patterns include ABCDEFGHIKLMNPQRSTVWXYZU; valid DNA characters for PHI-BLAST patterns include ACGT).

**MegaBLAST Tab**
- This program is optimized for aligning sequences that differ slightly and is up to 10 times faster than more common sequence similarity programs. It can be used to swiftly compare two large sets of sequences against each other. MegaBLAST permits searching with batches of ESTs or with large cDNA or genomic sequences.
- On the MegaBLAST Tab (Figure 35.5), you can set the percent identity used for the MegaBLAST search.

![Figure 35.4 PSI-PHI tab of the BLAST Search dialog box](image1)

![Figure 35.5 MegaBLAST tab of the BLAST Search dialog box](image2)
BLAST Options

To specify technical settings for the BLAST Search, choose View > Options, opening the following dialog box (Figure 35.6):

![Figure 35.6 Blast Options dialog box](image)

Complete your selections in the box. Press the Proxy Configuration button to set proxy server information (Figure 35.7).

![Figure 35.7 Internet Connection dialog box](image)

Make sure that the proxy settings are filled in correctly; specifying whether or not your connection to the Internet should be performed via a proxy server. Enter the address and port number of that server. If your proxy requires a password, check the Proxy requires authentication box and enter the user name and password information.

Click OK, OK, returning you to the BLAST Search Viewer.

Press the Submit button ( ) to launch the BLAST search.

BLAST Search Results

As soon as you have submitted your query, the search is numbered and listed in the Status Pane. This pane displays the following data about your search: data and time submitted, the status of the search (submitted, waiting, error, completed, etc.), and the identification number assigned your query. Completed searches are listed also showing the number of identified hits in the response.

**Note:** Depending on the stability and speed of your network connections, performance during search results downloading can vary. In the case of an unstable and/or slow network connection, large downloads may be aborted before completion. To avoid this problem, if you are downloading multiple molecules it is recommended that you download in several small batches instead of one large batch.
BLAST Viewer

Opening BLAST Viewer

The BLAST Viewer for reviewing BLAST search results can be opened by either of two methods:

1. Select **Start > Programs > Invitrogen > Vector NTI Suite 10 > BLAST Viewer**. If you choose this method, it opens as an empty viewer. You can open previously saved BLAST results by clicking on the **Open** button ( ) or by selecting **BLAST Results > Open**.

2. Double-click on a BLAST Search result entry in the BLAST Search dialog box. See the beginning of this chapter.

Elements of the BLAST Search Results Viewer

Blast Viewer displays the textual, sequence and graphical summary of the Blast search results (Figure 35.8). It consists of a menu bar, and two toolbars and three panes: Text, Analysis (which includes the Sequence Profile/Hit Distribution Pane, the Query-Hit Alignment Map Pane, and the Hit Map Pane) and Alignment Panes.

Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button. Many toolbar operations can also be launched from the menu bar or from a shortcut menu.

The BLAST Viewer is organized and managed like in the other display windows in Vector NTI Advance. Several BLAST Viewers can be opened in one workspace. You can navigate through the open display windows using the **Back** and **Forward** buttons ( ).

As in the other Vector NTI Advance applications, to apply a command in a given pane that pane must be active. The active pane can be toggled with the different panes buttons on the Pane Selection Toolbar ( ) or by clicking anywhere in the pane you wish to activate. The keyboard strokes are F6 (cycling clockwise through the panes) and SHIFT + F6 (cycling counter clockwise). **Note that in this BLAST Search Viewer, each of the graphs in the Analysis Pane is assigned a separate Pane button.**
Two management features of the display window are designed to maximize viewing flexibility:

1. Select View > Maximize Pane to expand the size of an activated pane to full screen size. Select View > Restore Layout to return to the original pane layout.

2. Select View > Edit Pane Layout to open the Pane Manager for modifying the arrangement of the panes in the workspace. For more details, refer to Chapter 3.

3. The Status bar shows the Hit Element Score from the Text Pane folder for the selected hit element.

In addition, the following pane manipulations are outlined in Chapter 3:

- Re-sizing the Text Pane with the split bars or arrow keys
- Scrolling the Text Pane

**BLAST Search Results**

**Text Pane**

The Text Pane of BLAST Viewer contains a description of the query molecule, including common database information, a Feature map and a list of hit molecules. The Text Pane is linked to analyses graphs in the Analysis Pane and the Alignment Pane.

The following Text Pane manipulations are outlined in Chapter 3.

- Opening or closing folders
- Opening the folder shortcut menu

The first line of the Text Pane lists the name of the query molecule. Double-click on the line or click on the at its left to open an Edit <molecule > dialog box. For details on editing the molecule, see Chapter 19.

The second entry in the Text Pane (Query Molecule) contains information about the query sequence such as the molecule length. It also contains the Search Parameters folder, which stores the parameters used in performing the BLAST search with this query molecule.

The Query Molecule section is followed by the Hits folder, which indicates the number of hits and contains the Hit molecules and related information. Click on the + at the left of the Hits folder to open it.

**Sorting Hits in the Text Pane**

To sort the hits in the Text Pane, right-click on the Hits folder to launch the shortcut menu from which the various sorting commands can be selected (Figure 35.9):
Deleting Hits from the Hits Folder
To delete an unwanted hit, right-click on the hit name and choose Delete from the shortcut menu.

Hit Molecule General Information
Notice that the hit molecules themselves are links. If you click on the link, it prompts the downloading of the molecule from the NCBI website, thus opening it in a Molecule Display window.

For each hit molecule, the following information is provided:
- Its ID code and the number of hit elements (regions of local similarity) in the molecule.
- A brief summary relating to the matching region of the hit molecule to the query sequence and the length of the matching region
- Score Folder(s) corresponding to the number of hit elements

Hit Element Description
Double-Click on a Score Folder for a particular hit element to open it. The folder displays the statistical results for the hit element such as its expectation value, identities, etc.

Selecting a Score Folder for a hit element locates all of the hit elements for that hit molecule in the Analysis Pane. However, just the hit element corresponding to the selected folder is displayed in a colored mode. As you click on the other Score Folders in the Text Pane, they are each highlighted in the Analysis Pane, in turn.

The statistical results in the Score Folder are described as follows:
- Score + (Bit Score): These indicate how close the identity of the match is to the query sequence.
- Expectation: This value reflects the likelihood that the similarity between the sequences would occur by chance when searching a database of a particular size. A zero or extremely low number suggests that the match is so perfect that it is extremely unlikely that the similarity would occur randomly. The E value describes random background noise exists for matches between sequences.
- In a BLAST search, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. For more detail about this score, see the BLAST parameters section at the beginning of this chapter.
- Identities: the ratio (and percentage) of matching residues in the hit elements. The numbers n/n refer to the number of identical residues out of the number of matches in the hit element. This is important to consider when determining the significance of this statistic. A high identity percentage may mean nothing if a low number of nucleotides is being compared.
- Positives: the ratio (and percentage) of similar residues in the hit elements
- Query: The position numbers in the query sequence matching that of the hit element are indicated by the start and end positions of the figures, such as 57[bp] – 93[bp]. The strand that corresponds to the hit element is also shown.
- Hit: The position numbers in the hit sequence matching the query sequence are indicated by the start and end positions of the figures, such as 42[bp] – 88[bp]. The strand that corresponds to the hit element is also shown.

Analysis Pane
The Analysis Pane, displaying the graphical summary of the BLAST Search results, consists of three panels: 1) the upper graph displays the Sequence Profile and the Hit Distribution, 2) the
center graph displays the Query-Hit Alignment, and 3) the lower graph displays the Hit Map graph.

Sequence Profile Graph (upper pane)

The Sequence Profile Graph depicts how well the particular regions of a query molecule are represented in a database it was searched against, allowing “unique” and “shared” regions of a query molecule to be easily identified. Each residue of the query sequence is assigned a value indicating the position-specific representation of this residue in a database.

The profile is calculated as follows: each residue is traced into all hit elements in which it is included. The corresponding residue substitution score from a hit element is multiplied by a logarithm of an expectation value of a particular hit element and the resulting value is summarized over all hit elements. Each step in the resulting graph represents the profile of each sequence residue.

If the database search utilized a translation of the original molecule as an intermediate step (tblastx and blastx), the sequence profile graph is generated for all possible translation frames.

Hit Distribution Graph (upper pane)

Each residue in the query sequence is traced to all hit elements in which it is included, and 100 percent identical. The number of hit elements per residue are tallied (summed up) and plotted across the molecule.

Query-Hit Alignment Graph (center pane)

Each element on the Hit Distribution Graph positions the local similarity regions on the query and hit molecules. The upper scale represents the query sequence while the lower scale represents the currently selected molecule in the hits list. The hit element corresponding to the currently selected Score Folder (Text Pane) is shown in color on the map, while the other hit elements for the molecule are grayed out. To select another element in the graph, you must return to the Text Pane and select a new Score Folder.

Hit Map Graph (lower pane)

The Hit Map graph provides a graphical overview of the hit sequences aligned to the query sequence. The X-axis on this map represents residues along the query molecule. Each bar drawn above the X-axis represents a molecule (or molecule fragment) that matches the query sequence. The position of each bar relative to the linear axis of the query allows users to see instantly the extent of the hits that align with single or multiple regions of the query.

Hit molecules are represented in decreasing order of similarity with the most similar hits shown at the top and the least similar shown at the bottom. If one hit molecule contains multiple hit elements, they are represented along the same Y-axis position, irrespective of similarity scores.

To view the Name, score, bits and E-value of a hit, pause the cursor over the hit element bar on the graph. Clicking on a hit element bar selects the hit element in the Text Pane. This also changes the bar’s color to indicate the selected hit and displays the alignment between the query and the hit sequences in the Query-Hit Alignment pane.

Zooming In and Out

Enlarging or reducing regions of the Analysis Pane is possible using the Zoom buttons. Click in the pane for the graph to be resized. If you Zoom In sufficiently, you can see the individual residues on the X axis. For more detailed about using the Zoom buttons, refer to Chapter 3.

To view a specific selected region in one of the graphs, select Zoom to Selection from the shortcut menu. To return to the normal mode, click on the **Fit to Window** button ( ).
Note: In the Analysis Pane, only the upper and lower graphs can be resized.

Selecting Regions on the Analysis Pane

Regions of the graphs can be selected in the Analysis Pane by holding down the left mouse but-
tton then dragging through the sequence. Selected regions are displayed concurrently on other
graphs and the Alignment Pane, (only if the Score Folder for the hit element corresponding to
the selected region is selected in the Text Pane.)

Hits listed in the Text Pane that correspond with the selection in the Analysis graphs can be
highlighted using the Select Hits function. To use the Select Hits function, highlight the area of
interest on the Analysis Pane graphs and choose Edit > Select Hits from the menu. The Select
Hits dialog box appears (Figure 35.10):

![Select Hits dialog box](image)

There are four choices for mapping the selection to the hits in the Text Pane:

- **Intersecting with selection** – hits that contain any part of the selected area.
- **Contained within selection** – the entire length of the hit matches within the selection.
- **Containing entire selection** – hits contain the entire selection and, possibly, more.
- **All** – hits with any of the above matching are found.

Select the radio button for the intersection option of your choice and press the OK button. Hits
corresponding in the Text Pane corresponding to the selection in the Analysis Pane and the
selected intersection option are highlighted.

Viewing/Changing Parameters of the Analysis Graphs

The graphs in the Analysis Pane display the values averaged in a window of a specific length
sliding along the x-axis of the graph by a specific step.

To view and modify the parameters of the Sequence Profile or Hit Distribution plots, activate the
upper graph pane by clicking on its Pane button, or click in the upper pane. Select Plot Setup
from its associated shortcut menu, opening the Plot Setup dialog box:

- **Analysis tab**: Count Window Size is the length of the sliding window; the larger the
  Count Window Size, the smoother the curve. Step Size is the width of each step on the
  plot.
- **Sequence Profile/Hit Distribution tabs**: Modify the plot color and select the graph dis-
  play mode from Bar, Boolean, Polygon, or Linear.

To view and modify the Hit Map graph properties, activate the Hit Map pane by clicking on it or
by pressing its activation button on the toolbar. Right-click on the Hit Map graph pane and
choose Properties from the shortcut menu or choose Edit > Properties from the menu to
launch the Hit Map Properties dialog box. In the Hit Map properties dialog box, you can change
the display color for hits and the selected hit and adjust the size of the graphic hits display.

Alignment Pane

The Alignment Pane displays the alignment of the hit element selected in the Text Pane and the
 corresponding region on the query sequence. The query molecule, the currently selected hit ele-
ment and the resulting consensus sequences are shown.
Moving Around the Alignment
The Alignment Pane has its own horizontal scrollbar, if the sequences are sufficiently long to warrant them. The sequence names remain stationary at the left in the Alignment Pane no matter how the sequences are repositioned horizontally. The position number of the left-most residue visible in each sequence is displayed immediately following each sequence name.

Selection Techniques
Select molecules in the Alignment Pane by clicking on their names.
To select regions of the nucleotide or amino acid sequence alignments, use the click and drag technique with the mouse. Selection applies only on the sequence of the currently selected molecule. The selection is also shown on the graphs in the Analysis Pane.

Alignment Properties
To change a default set of alignment display parameters, from the Alignment Pane, select Properties from the shortcut menu or click on the Properties button. The Alignment Display Setup dialog box opens with three property tabs: Consensus Calculation, Similarity Tables, and Color Setup.

Consensus Calculation
A consensus sequence is a theoretical representative nucleotide sequence in which each nucleotide represents either the residue seen most frequently at that same site in aligned sequences, or is selected by other criteria. The Consensus Calculation tab specifies how the consensus sequence, displayed as the bottom sequence in the Alignment Pane, is calculated in BLAST Viewer.

On the Consensus Calculation Tab, you can select among the following parameters:
- Consider only identical residues
- Ignore gaps in consensus calculation
- Specify a sequence to use as the consensus sequence. In this case, select one of the sequences listed (those used in the alignment).
- Residue fraction

If the Use sequence as consensus checkbox is checked, the Consensus will not be calculated but the sequence selected in the listbox below the checkbox will be used as the consensus.

Similarity Tables
The Similarity Table allows you to review or to define residue similarity information and values necessary for calculation of the alignment quality profile. Each residue pair can be defined as being strongly or weakly similar to each other. The table is editable only for amino acid residues.

Color Setup
The Color Setup tab allows you to review or modify the current color setup scheme for the Alignment Pane:

<table>
<thead>
<tr>
<th>Color Scheme</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>black on window default</td>
<td>non-homologous residues (Non-similar)</td>
</tr>
<tr>
<td>color</td>
<td></td>
</tr>
<tr>
<td>blue on cyan</td>
<td>consensus residue derived from a block of similar</td>
</tr>
<tr>
<td></td>
<td>residues at a given position (Conservative)</td>
</tr>
</tbody>
</table>

Table 35.2 Color Setup tab interpretations of color scheme
Saving BLAST Search Results

To save BLAST search results into the database for future retrieval, select **Blast Results > Save As** or click the **Save As** button ( ). Select one of the three tabs in the Save As dialog box (Figure 35.11), depending on your preferred destination. The results can be saved in the Vector NTI database or as a BLAST result file.

Select the subset or folder (and file format for Results File) and enter a name for the results (Figure 35.12).

Click **OK** to save the BLAST results as a new database object, associated with the query molecule. You can open search results for the query molecule from the Molecule Viewer window.

---

### Table 35.2  Color Setup tab interpretations of color scheme (continued)

<table>
<thead>
<tr>
<th>Color Scheme</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>black on green</td>
<td>black on green consensus residue derived from a single conservative residue at a given position (Block of Similar)</td>
</tr>
<tr>
<td>red on yellow</td>
<td>red on yellow consensus residue derived from completely conservative residues at a given position (Identical)</td>
</tr>
<tr>
<td>green on window default color</td>
<td>green on window default color residue weakly homologous to consensus residue at given position (Weakly similar)</td>
</tr>
</tbody>
</table>

---
**Saving or Opening Selected Hits**

To save individual BLAST hits into the Vector NTI database, select the desired search hits and choose **Save Selected Hits** from the shortcut menu, opened by a right click on the molecule name. Hits will be saved to corresponding molecules and can be viewed later in the Molecule Viewer.

To open individual BLAST hits into Vector NTI, select the desired search hits and choose **Open Selected Hits** from the shortcut menu, opened by a right click on the molecule name. The molecule(s) can then be saved to the VNTI Database or to files on the hard drive.

**Exporting or Opening Query Molecule with Hits Displayed as Features**

You can open the query molecule that was used for the BLAST search in Vector NTI with the BLAST hits displayed on the molecule as features or export it to a .gff file. In both cases, hits are displayed graphically on the query molecule as the feature type of your choice.

When you execute the **Open Query Molecule with Hits as Features** command (under Edit in the Main Menu), the molecule opens in a Vector NTI Molecule Viewing window. You can then save the annotated query molecule to the VNTI Database.

When you export a query molecule with hits displayed as feature into a .gff file, you use the **Export Query Molecule with Hits as Features into GFF file** command to save the .gff file to an external location (hard drive, network, floppy, etc.). You can open a .gff file in a Vector NTI Molecule Viewing window and save it to the Vector NTI Database, if desired. .gff files provide a method for sharing query-blast hit information in Molecule format with other Vector NTI users.

**Opening Query Molecule with Hits as Features**

To open the query molecule in VNTI with the hit sequences displayed as features, choose the **Edit > Open Query Molecule with Hits as Features** command from the menu. The Choose Feature dialog box appears allowing you to choose the feature type you want to use. Choose the desired feature type, check the box to save existing query molecule features, if desired, and click the **OK** button. The query molecule opens with the BLAST hits displayed as the chosen feature type.

In the Text Pane of the new molecule, query hits are listed in the chosen feature type folder. Hit feature names follow the convention “Hit Molecule [gi# |gb# |EMBL# |Reference#] Information; Hit #, Element #” where

- Hit Molecule Information lists available reference numbers for the Hit Molecule
- Hit # is the hit molecule number according to its original listing in the BLAST Viewer
- Element # is the hit element number of the hit molecule.

The new molecule can then be saved to the VNTI Database or to a molecule file on the hard drive.

**Exporting Query Molecule with Hits as Features**

To export the query molecule with the hit sequences displayed as features, choose the **Edit > Export Query Molecule with Hits as Features into GFF file** command from the menu. The Save As dialog box appears, allowing you to select the storage location for the file. Next, the Choose Feature dialog box appears allowing you to choose the feature type you want to use. Choose the desired feature type, check the box to save existing query molecule features, if desired, and click the **OK** button. The query molecule with the BLAST hits displayed as the chosen feature type is saved to a .gff file in the designated location.
Creating a New Database Subset for BLAST Search Results

To create a new database subset for storing BLAST search results, in the BLAST Viewer, click the **Database** button ( ), opening Vector NTI Advance’s Database Explorer. The Explorer opens to the BLAST Results table. To create a new subset for storing BLAST Search results, follow the instructions given in Chapter 17.

Other BLAST Viewer Features

The following features can be performed in BLAST Viewer. Details are given in Chapter 3.

- Copying selected molecule IDs to the clipboard
- Exporting data to external programs
- Printing from the BLAST Viewer
Introduction

AlignX Blocks is a program for locating, analyzing, and editing blocks of localized sequence similarity among multiple protein sequences and linking them into a composite multiple alignment.

The following definitions are important in the AlignX Blocks application:

- A Block is a set of sequence segments of the same length taken from several protein sequences and aligned without gaps.
- A Local Alignment is an alignment that includes only the most similar regions (even fairly small) of the sequences under consideration. Such regions may span several protein sequences.

A Blocks project, including molecules, settings and blocks generated in AlignX Blocks can be saved as a file. This enables you to restore the project later to continue working with it.

Opening AlignX Blocks:

- To open AlignX Blocks, select Start > Programs > Invitrogen > Vector NTI Suite 10 > AlignX Blocks. You can also open AlignX Blocks from the Vector NTI workspace, Database Explorer or other Vector NTI applications (such as AlignX or BioAnnotator) under the Align menu option.

Features of the AlignX Blocks Workspace

When AlignX Blocks is first opened, the display window is empty. As molecules are added to an AlignX Blocks project, they are listed in the Text Pane. After performing a Block Search on a group of molecules selected in the Text Pane, blocks results appear in the other panes.

You may want to open a demo project into the workspace before you review its features. Select Project > Open or click the Open Project button ( ). Find the Vector NTI Suite folder, probably located in Program Files\Vector NTI\Demo Project, and open the Protein.bpr file.
The AlignX Blocks interface (Figure 36.1) consists of a menu bar, three toolbars, an AlignX Blocks display window, with shortcut menus specific for each pane. The AlignX Display window includes four panes.

AlignX Blocks has three toolbars: a Main Toolbar, the Pane Selection Toolbar and a Pane Toolbar. Most toolbar buttons have associated tool tips that explain the button function; hover the cursor over the button to display the tool tip for that button. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.

The AlignX Display window is divided into four panes: a Text Pane, the Block List Pane, an Alignment Pane and a Schematic Pane. All panes are separated by split bars that allow resizing of the panes. Scroll bars appear when the contents of a pane exceed the viewing area, facilitating easier viewing of pane contents.

As in the other Vector NTI Advance applications, to apply a command from the main menu in a given pane (Print, Print Preview, Camera), that pane must be active. The active pane can be toggled with the Switch Panes ( , , , ) buttons in the Pane Selection Toolbar or by clicking anywhere in the pane you want to activate.

You can maximize any active pane to fill the entire display space by selecting View > Maximize Pane. Select View > Restore Layout to restore the original pane size.

You can modify the pane arrangement in Pane Manager. Select View > Edit View Layout to open the Pane Manager. For more details, refer to Chapter 3.

Text Pane

To activate the Text Pane, click on the Text Pane button ( ) on the Pane Selection Toolbar.
The Text Pane contains a thorough description of the molecules included in the AlignX Blocks project. The information is organized in folders, named by molecule. Each top level (molecule) folder contains the following subfolders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>General description</td>
<td>Molecule type, length, etc.</td>
</tr>
<tr>
<td>Comments</td>
<td>Arbitrary text of any length associated with the molecule</td>
</tr>
<tr>
<td>Standard fields</td>
<td>GenBank/SWISS-PROT-like fields: keywords, division, original accession numbers, etc.</td>
</tr>
<tr>
<td>References</td>
<td>Bibliographic references (in GenBank/SWISS-PROT format)</td>
</tr>
<tr>
<td>Proprietary fields</td>
<td>User defined fields (for molecules which came from Vector NTI)</td>
</tr>
<tr>
<td>Analysis</td>
<td>Results of common physiochemical analyses</td>
</tr>
<tr>
<td>Feature Map</td>
<td>List of molecule features</td>
</tr>
</tbody>
</table>

Table 36.1 AlignX Display window Text Pane folders

Folders and sub-folders can be opened with a double-click or by clicking on the + to the left of the folder name.

For more information regarding the following Text Pane manipulations, refer to Chapter 3.

- Using the scrolling bars
- Expanding and collapsing folders
- Copying pane contents to the clipboard

**Finding Signals on Schematic and Alignment Panes**

Signals or features listed in the Feature Map subfolder in the Text Pane can be found in Schematic and Alignment Panes. This feature is only enabled if the molecules are involved in the blocks located in the Blocks Search, indicated by a blue arrow. To find the positions of the signal in these panes, select the feature line, indicated by the tag icon ( ), and click the **Find** button ( ) from the Text Pane toolbar or select **Edit > Find**. The **Find command is disabled if the selected line is not a signal with non-zero boundaries.**

**Block List Pane**

All of the Blocks, found in a Block Search or manually defined, are shown in the Block List Pane (Figure 36.2). To activate this pane, click within the pane or click on its button ( ).

![Figure 36.2 AlignX Display window Block List Pane](image)
Each line in the pane represents one block. The following columns (fields) are defined for the block:

- **Color representation**: A color is assigned to each block automatically from the list of available colors (defined in Project properties). The first block in the list gets the first color, second block – the second color, etc. If there are more blocks than colors the same colors are used again.
- **Name**: Initially, blocks found in a search have no name. Manually defined blocks are automatically named “ud1”, “ud2”, etc. You can name or rename a block in the Block Edit dialog.
- **m**: Number of sequences in a block
- **Length**: Length of the block, that is, the length of the sequence segments in the block. Like segments for all the sequences in a block have the same length.
- **MP-score**: Mean Pairwise Score, calculated by dividing the SP-score of a block by the number of possible pairwise combinations of component sequence segments.
- **SP-score**: Sum of the Pairs Score, composed of several sequence segments. Scores are calculated for all possible segment pairs. The sum of all of the scores is the SP-score.
- **Significance**: Biological significance (see "Methods for Assessing the Statistical Significance of Molecular Sequence Features by Using General Scoring Schemes" by Samuel Karlin and Stephen F. Altschul. 1990).
- **Offset**: Minimal offset of the segment in the block.

**Block Selection and Highlighting**

There are two methods of selection in the Block List Pane, each offering different functionality. Both types of selection work independently and do not affect each other.

1. **Single selection**: Because this selection enables actions within a block, only one block at a time can be selected with this method. Click on the block line outside of the colored circle area. The highlighted selected block is shown concurrently in the Schematic and Alignment Panes. To deselect the block, click on the Schematic or Alignment Panes.

2. **Multiple selection**: This selection method allows no actions within blocks, but is only for viewing them. Select one block by clicking within the colored circle area. Press SHIFT + CLICK within circles to select a range of several blocks. Press CTRL + CLICK within circles to select more than one non-contiguous block. To deselect the blocks, repeat the clicks on the selected circles.

Blocks selected in this way are indicated by the black dot in the center of the colored circle in the Block List Pane. The colored blocks on the Schematic and Alignment Panes correspond to the selected circles, distinguishing one from the other. Mouse clicking within Schematic and Alignment Panes does NOT remove this kind of selection.

**Sorting Blocks in the Block List Pane**

Initially blocks are sorted by the number of sequences in the block (m-column). The blocks order in the Block List Pane can be changed, however, by sorting. Click on a column header to sort; repeat the click to reverse the sorting order. Sorting is unaffected by selection of one or more blocks.

**Removing Blocks**

To delete a selected block, choose View > Delete Block or the corresponding command from the Shortcut menu. You can confirm or cancel the deletion.
Schematic Pane

This pane is designed for a schematic representation of blocks to facilitate analyses of their position and relationship. Click in the pane or on the Schematic Pane button ( ) to activate this pane.

Molecules included in the Block Search operation as well as manually added from the Text Pane are displayed schematically as stretched rectangles (Figure 36.3). A schematic representation of one sequence occupies each row, with the name of its sequence displayed to the left. The upper row of the pane is a ruler. Both molecule name and its sequence can be scrolled in the vertical direction. There is no horizontal scrolling bar for this pane because the display is calculated such that the longest rectangle fits the pane width.

You can change the order of the molecules in the pane by dragging a molecule with the DRAG + DROP technique. The order of the molecules in the Alignment Pane reflects the changes.

To select molecule(s) in the Schematic Pane, click on the name(s). Selections are highlighted concurrently in the Alignment Pane. To select a region in the Schematic Pane, click within the pane and drag a rectangle.

**Note:** In this pane, you can only select regions; you cannot select blocks.

Alignment Pane

The Alignment Pane (Figure 36.4) is designed for viewing and analyzing sequential information. To activate it, click within pane or click on the Alignment Pane button ( ) of the Pane selection toolbar.

Molecules included in the Block Search operation as well as those manually added from the Text Pane are displayed as sequences in the Alignment Pane. The name of the molecule and current position within the sequence appears to the left of its sequence. The upper row of the Alignment Pane contains a ruler with a scale of residues.

The Alignment Pane has its own vertical and horizontal scrollbars. Both a molecule name and its sequence can be scrolled in the vertical direction. When scrolled horizontally, the sequence names remain stationary at the left of the Alignment Pane. The number to the left of the sequence shows its current position.
Dragging the molecule (up or down) by the name and dropping it on the desired place can change the order of the molecules in the pane. The order of the molecules in the Schematic Pane is changed to reflect the order in the Alignment Pane.

**Selection in the Alignment Pane**

To select a molecule(s) in the Alignment Pane, click on its name. This selection is selected concurrently in the Schematic Pane as well.

To select a region in the Alignment Pane, click and drag within the pane. The region inside the tracking dotted rectangle is selected except for the “gapped” (or empty) areas. To combine a new selection with a previous selection, press the CTRL key while starting the new selection. To modify a selection, drag the edge of the selection to the right or left. Selections are displayed concurrently in the Alignment and Schematic Panes.

**AlignX Blocks Projects**

A set of molecules, options and blocks opened and being worked on is called a AlignX Blocks project. AlignX Blocks accepts following file formats: GenPept, SWISS-PROT, FASTA and Vector NTI archives.

**Opening an Existing Project**

When AlignX Blocks is launched, a new, empty project is created. To open an existing project, choose **Project > Open** or click on the **Open Project** button ( ) on the Main toolbar. Select the desired project name in the File Open dialog box. *AlignX Blocks projects have a “.bpr” file extension by default.* Select the desired file and click **Open**. If the file is a valid AlignX Blocks project, information about molecules, properties and blocks (if any) will appear in appropriate panes. An alternative way to open an existing project is to drag and drop the project file on the AlignX Blocks workspace.

**Creating a New Project**

To create a new, empty project, select **Project > New**. If appropriate, you will be prompted to save any changes in the currently opened project first. If you select **Cancel**, work may continue on the opened project. If you select **Yes** to save changes or **No** to disregard changes, the current project is closed and the workspace emptied.

**Adding Molecules to the Project**

To add molecules to a current project, click the **Add Files** button ( ) or select **Project > Add Files**. *Only protein molecules in acceptable format can be added to the project.* Molecule(s) added to the project are listed in the Text Pane.

You can also add molecules to the project using the drag and drop technique if they are of an appropriate file format and molecule type. If the program fails to define a file’s format, it will be suggested the file be imported as a raw sequence. Imported sequences appear in the Text Pane with the names: NONAME, NONAME#2, NONAME#3, etc.

**Block Search**

To search for blocks, select two or more molecules in the Text Pane and click the **Search for Blocks** button ( ) or choose **Blocks > Search for Blocks**. The Block Search progress dialog box opens, letting you monitor the search progress. Press the **Cancel** button to interrupt the current search. Blocks located in the search (if any) are shown in the Block List Pane, sequences of the selected molecules are displayed in the Alignment Pane and schematic representations of the sequences in the Schematic Pane.
Adding Sequences to Block Sequence Space.

The term Block Sequence Space (BSS) designates all the sequences that are displayed in the Schematic and Alignment Panes. Usually sequences are added to BSS as a result of a Block Search. Selected sequences appear in both Schematic and Alignment Panes regardless of whether or not any block includes this particular sequence. To add a sequence currently not present in a BSS, select Add Sequence from the shortcut menu opened from the Text Pane. If a sequence already is included in a BSS, this menu item is disabled. A sequence added using the Add Sequence command is also marked with blue arrow in the Text Pane.

Defining a Block

In AlignX Blocks, you can search for blocks automatically or define blocks manually. To define a block, select a region within the Alignment or Schematic Pane and click the Define Block button ( ), select Blocks > Define Block or the corresponding command from the Shortcut menu. All the segments of the newly defined block will have an offset number defined by the selection, and the length equal to the minimal segment length. All previous selections are overwritten and the new block is selected in all three panes. Manually defined blocks are named automatically as “ud1”, “ud2”, etc., but they can be renamed later.

Linking and Unlinking

Linking means aligning the currently selected sequence segments and creating “links” between the residues appearing in each effected column. When residues are linked, they remain in the same column despite any shifts that occur elsewhere in the alignment. Linked residues are shown in upper case in the Alignment Pane and as wider rectangles in the Schematic Pane.

To link, first select sequence segments within the Alignment or Schematic Panes, or select a block in the Block List pane. Click the Link Selected Area button ( ) or select Blocks > Link or the corresponding option on the shortcut menu.

To unlink selected sequence segment(s), click the Unlink Selected Area button ( ) or select Blocks > Unlink or the corresponding command on the shortcut menu. Unlinking selected segment(s) can cause segments in other sequences to also become unlinked.

To unlink all segments of all sequences, select Blocks > Unlink All.

When you perform Link functions, the previous state is saved in the history list.

The Undo Link Operation button ( ) and Redo Link Operation button ( ) buttons on the Main toolbar allow navigation through the history list. Information about links is saved in saved in an AlignX Blocks project and applied to Alignment Pane next time you open the project.

Color Setup

Colors for displaying linked and unlinked areas in the Alignment Pane can be changed in the Link Color dialog box (Figure 36.5). To modify the Link Color, select View > Color Setup or the corresponding command from the shortcut menu, opening the Link Color dialog box.

Figure 36.5 Link Color dialog box
This dialog box lets you select the background sequence color (Text color) and Linked text color. Select the appropriate color(s) and click **Apply** to preview the selections. Click **Close** to apply the changes and close the menu.

For more information regarding linking and unlinking, see the previous section.

Block Editing

To edit a block, click the **Edit Selected Block(s)** button ( ), or choose **Blocks > Edit Block** or the corresponding command from the shortcut menu, launched by right clicking on the desired block or double click on the desired block. This opens the Edit Block dialog box (Figure 36.6).

![Edit Block dialog box](image)

**Figure 36.6 Edit Block dialog box**

**In the Block Editor:**

1. Remove sequences from a block by clicking on their number/name (at the far left).
2. Narrow or widen a block by dragging the line separating it from the rest of the sequence. All changes in block statistics are immediately displayed in the current line (in the section below the Alignment Pane).
3. Name or rename a block in the block name text box.
4. Change a block’s color in the block color dropdown menu.
5. Press the **Restore to original** button to undo all changes.
6. After changes are saved, the Block List, the Schematic and Alignment Panes are updated.
7. The block statistics in Block Editor are described as follows:
   - **m:** number of sequences in a block
   - **Length:** Length of the block, that is, the length of the sequence segments in the block. Like segments for all the sequences in a block have the same length.
   - **MP-score:** Mean Pairwise Score. The MP-score is calculated by dividing the SP-score of a block by the number of possible pairwise combinations of component sequence segments.
   - **SP-score:** Sum of the Pairs Score, composed of several sequence segments. Scores are calculated for all possible segment pairs. The sum of all of the scores is the SP-score.
   - **Significant?:** Significance of a block. Standard P-value (Karlin, Altschul, 1990)
Deleting a Molecule

To delete a selected molecule from the project, click the **Delete from Project** button ( ) or choose the **Delete < Molecule> From Project** from the shortcut menu.

If a selected molecule participates in one or more blocks, you will be asked to verify the deletion. *All blocks it participates in will be dismissed.* If you confirm the deletion, the specified molecule and all related blocks are removed from all panes.

**AlignX Blocks Parameters**

To review or modify AlignX parameters setup, click on the **AlignX Blocks Setup** button ( ) or select **Blocks > AlignX Blocks Setup**. This opens the AlignX Blocks Parameters dialog box (Figure 36.7).

On the three tabs, review or modify the following parameters:

**Parameters Tab:**

- **Pairwise score cutoff:** the minimum segment pair score needed to mark a diagonal
- **Min. sequences per block:** the minimum length of sequence that block must span in order to be reported
- **Expectation value limit:** the maximum threshold value required to make block significant

![Figure 36.7 AlignX Blocks Parameters dialog box (Parameters tab)](image)

**Blocks’ Color Tab:**

To change Block color(s), select the appropriate color from the palette then click **OK**.

There are two color listboxes on this tab: **Block colors** selected for this project and **Available Colors** (Figure 36.8). Double click on any color from the Available Colors list to add it to the
Blocks Colors list. To remove a color from the current block double click on the color in the Block colors list.

![Figure 36.8 Blocks' Colors tab on the AlignX Blocks Parameters dialog box](image)

Score Matrix Tab

The Score Matrix tab (Figure 36.9) displays the matrix used to generate the current multiple alignment in AlignX Blocks. From this tab, you can also open other matrices for review or for application to the Blocks. Editing a matrix takes place in the Matrix Editor, accessible from AlignX and AlignX Blocks.

When a project is created, a residue substitution matrix from the corresponding protein template is used. If a template file is not found, the default matrix is used (blosum62mt2 for protein) or enter the name of a new score matrix in the Name text box. To review other matrices, click the Select Matrix button. From the Matrices file, select any other matrix to be loaded on the Score Matrix tab.

![Figure 36.9 Score Matrix tab on the AlignX Blocks Parameters dialog box](image)

For further details regarding customizing or editing matrices, refer to Chapter 39.

Templates

Templates are the files that contain all the properties for the AlignX Blocks program. An AlignX Blocks project is created using default template file. You can also save current settings of the AlignX Blocks project in a template file or apply settings from existing template to the open project.

To save settings in a template choose Edit > Setup > Save to File. Enter or select the file to which settings will be saved.
To apply settings from existing template to the project that is already open choose Edit > Setup > Load from File. All the properties are applied after researching blocks.

**Other AlignX Blocks Features**

The following features can be performed in AlignX Blocks. Details are given in Chapter 3.

- Copying AlignX Blocks data to the clipboard
- Printing from AlignX Blocks
- Exporting molecules from AlignX Blocks
- Integrating AlignX Blocks with other Vector NTI applications and third-party tools on the Internet.
Introduction

The three-dimensional representation of a protein is directly linked to the linear sequence of the protein. Therefore, the structural information about the protein regions derived from a BLAST search, multiple alignment, mutation analysis, BLOCKS or ProSite database search, etc. can be easily visualized on the protein structure. Similarly, 3D features of proteins, such as protein-protein interaction sites or small molecule binding sites, can be mapped on the linear protein sequence to elucidate functional properties of particular protein regions.

Resolved three-dimensional structures of biological macromolecules are stored as public domain files in the Protein Data Bank (PDB). 3D-Mol is a Vector NTI Advance application used for visualization and analysis of three-dimensional information derived from PDB structures or structures created by the user. The 3D viewer is designed to simplify viewing and analysis of tertiary structures of proteins, nucleic acids and small organic and inorganic molecules.

3D-Mol Supported File Formats

3D-Mol accepts the following file formats:
- 3D Molecule Viewer Project File (.p3d)
- Brookhaven PDB File (.pdb, .ent)
- Version 1.0 M3D File (.m3d)
- Compound File (.m3dc, .m3dz)

The viewer currently allows displayed molecules to be exported in PDB and M3D formats.

Downloading 3D Structure Files

You can search for PDB files in the public database using the Vector NTI Online Resources. In Vector NTI, choose Tools > Online Resources > Entrez Search System at NCBI tool or the software application, PubMed/Entrez Search. For the database type, enter “structure.” When you receive the result(s), copy the file(s) as a document and save it for opening in 3D-Mol, drag
and drop it into this application, or copy it in PubMed/Entrez Search and paste it into the 3D-Mol workspace.

For more information on the PubMed/Entrez Search, see Chapter 39 of this User’s Manual.

Opening 3D-Mol

To open 3D-Mol, choose Start > Programs > Invitrogen > Vector NTI Suite 10 > 3D Molecule Viewer. When 3D-Mol is first opened, the display window is empty. As molecules are added to a 3D-Mol project, they are listed in the Project Pane.

To open 3-dimensional structure files in 3D-Mol, press the Open button ( ) or select File > Open. In the Open dialog box, navigate to the structure file, highlight it and press the Open button.

Elements of 3D-Mol

The 3D-Mol interface (Figure 37.1) consists of a menu bar, three toolbars, and a 3-D Mol display window, with shortcut menus specific for each pane. The 3D-Mol Main toolbar contains tools for major program functions, such as opening and saving projects, adding molecules to the project, exporting molecule data in various formats, and optimizing the 3D-Mol window view. The Select and 3D toolbars contain tools for manipulating the graphic representation of the structure. The Pane toolbars and drop-down menus have different modes, depending on the active pane.

Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button. Many toolbar commands can also be launched from the menu bar or from a shortcut menu.

The 3D-Mol display window consists of four main panes: the Project Pane that displays information about the project, the Structure Pane that displays information about molecules that are opened in 3D-Mol, the Graphics Pane used for the visual display of 3-D structural data, and the Sequence Pane with separate tabs for viewing each sequence chain in the protein and/or nucleic acid structure being viewed.

As in the other Vector NTI Suite applications, to apply a command from the main menu in a given pane, that pane must be active. A pane can be activated by clicking anywhere in the pane or by choosing Window > <pane name> from the menu.
All panes are separated by split bars, which allow re-sizing of the panes. Scroll bars appear when the contents of a pane exceed the viewing area, facilitating easier viewing of pane contents.

You can add or remove a pane from the 3D-Mol window view using the Pane buttons in the Pane Selection toolbar. Under the Window menu option, you can also select options that allow you to show or hide the various toolbars or the status bar.

**Project Pane**

Click on the Project Pane to activate it. The Project Pane contains a list of the structures in the project organized in folders according to structure file type (Figure 37.2).

![Figure 37.2 3D-Mol Project Pane](image)

Double-click on a structure file name in the Project Pane or highlight the structure and choose **View > Activate Molecule** to display the structure in the 3D-Mol workspace. The currently displayed structure shows a check mark by its bolded name in the Project Pane (Figure 37.2).

**Viewing Structure File Properties**

Highlight a structure in the list and choose **View > Properties** or right-click on the structure and choose **Properties** from the shortcut menu to launch the File Properties dialog box for that structure (Figure 37.3):

![Figure 37.3 File Properties dialog box](image)

- **File tab** – Contains physical data about the structure, such as hard drive location, size and modification date.
- **Header tab** – Contains information from the structure file header.
- **Remarks** – Contains remarks from the structure file.

**Creating Custom Structures (NewEntries)**

You can create user-defined structure files in a 3D-Mol project. This is accomplished using the NewEntry feature of 3D-Mol.

**Adding NewEntries to the Project**

You can add a user-defined structure by activating the Project Pane and choosing **File > New Editable Entry** from the menu. This command adds a new folder called *Edits* to the Project Pane, if one does not already exist, along with a new structure called *NewEntry<entry #>* (Fig-
Subsequent new entries created are named with the next consecutive entry number (i.e. NewEntry1, NewEntry2, NewEntry3, etc.).

Adding Content to NewEntries

Unlike imported structure files, NewEntries are editable, allowing you to create your own custom structure files. Commands available on the Structure Pane and Graphics Pane shortcut menus provide the means for copying chains from an imported or user-defined (NewEntry) structure file and pasting them into a NewEntry to create a custom structure file. Once you have added content to a NewEntry, an asterisk appears by its name in the Project Pane (Figure 37.4) and in the 3D-Mol title bar when the modified NewEntry is selected for viewing:

![Figure 37.4 Project Pane showing NewEntry with unsaved added content](image)

To copy a chain, right-click on the chain in the Structure Pane and choose **Copy Chain** from the shortcut menu. To paste the copied chain into a NewEntry file, double-click on the NewEntry in the Project Pane to activate it, right-click in the Graphics Pane and choose **Paste Chain** from the shortcut menu or activate the Graphics Pane and choose **View > Paste Chain** from the menu. Once the chain is pasted, you can move it in the Graphics Pane using the techniques described in the *Moving Molecules* section below.

Once you have created a custom structure, you can save it to a structure file by choosing **File > Export Entry** from the menu (see also *Exporting Structures to a File*, page 624). Once you have exported the NewEntry to a file, the name appears in the Project Pane with the file type extension (Figure 37.5):

![Figure 37.5 Project Pane with two NewEntries that have been save to files](image)

Working with NewEntries

You can manipulate custom structure files in 3D-Mol in the same way you can downloaded structure files.

**Removing a Structure from the Project**

To remove a structure, highlight the structure in the Project Pane and choose **View > Remove Molecule** from the menu or right-click on the structure and choose **Remove Molecule** from the shortcut menu.
Structure Pane

Click on the Structure Pane to activate it. The Structure Pane contains a description of the structure opened in 3D-Mol (Figure 37.6).

![Figure 37.6 3D-Mol Structure Pane](image)

Structure information is organized in folders, based on molecule and atom data. Each top level (molecule) folder contains the following subfolders:

<table>
<thead>
<tr>
<th>Folder</th>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecules</td>
<td>Contains data about the structure such as number of chains, chain types and names, functional groups, etc.</td>
</tr>
<tr>
<td>Atoms</td>
<td>Characterizes atoms in the structure according to Element, ID, Amino Acid and Secondary Structure.</td>
</tr>
</tbody>
</table>

*Table 37.1 The 3D-Mol display window Structure Pane folders*

Folders and sub-folders can be opened with a double-click or by clicking on the + to the left of the folder name.

For more information regarding the following Text Pane manipulations, refer to Chapter 3.

- Expanding and collapsing folders
- Copying pane contents to the clipboard

**Finding and Marking Signals on Graphics and Sequence Panes**

Signals or features listed in molecule subfolders in the Structure Pane can be marked and displayed in the Graphics and Sequence Panes. Select the feature line in a Structure Pane folder, and click the Mark Selection button ( ) or select Mark Selection from the feature’s shortcut menu. For the feature(s) to be shown in the 3-D structure, click the Show Selection button ( ). To leave the item(s) selected, but hidden, click the Hide Selection button ( ). To deselect the item, click the Unmark Selection button ( ) or select View > Unmark Selection. You can mark multiple features, residues or atoms if you wish. To clear all marked regions, click the Unmark All button ( ).

By default, marked signals are painted in the Graphics Pane in the color designated in the 3D Options dialog box (see Sequence Pane Viewing Options, page 620). However, marked atoms can be left marked, but unpainted if you want to temporarily remove their marks from view. To paint/unpaint marked atoms, click the Paint Marked Atoms button ( ) on the 3D-Mol toolbar.
Mark in Radius

When atoms, rotamers and monomers are marked from the Structure Pane, all like species within the set mark radius are also marked. The Mark in Radius command allows you to change the marking radius.

To change the marking radius, right-click in the Structure Pane and select Mark in Radius > Set Radius (current set radius) from the shortcut menu. In the Set Selection Radius dialog box that opens, enter a radius between 1 and 20 angstroms and click the OK button. Subsequent atoms, rotamers and monomers marked from the Structure Pane will mark all like species within the new designated mark radius.

Graphics Pane

Click on the Graphics Pane to activate it. The Graphics Pane shows a visual display of the three-dimensional molecule structural information. Graphics Pane tools enable rotation of the molecule(s) displayed there.

To select features of a molecule, mark the part of the molecule using the selection tools. All selections except the selection of individual atoms are displayed concurrently on the Sequence and Alignment Panes.

Point to any atom in the Graphics Pane to view a tool tip with information about the atom, such as chain and structure it belongs to.

Moving Molecules

The graphic representation of molecule(s) in the Graphics Pane can be moved, rotated and zoomed in and out. There are several methods for moving molecule graphics:

- Press and hold left mouse button or activate the Graphics Pane and use the keyboard arrow keys to rotate molecule(s) around vertical and horizontal axes lying in the plane of the screen.
- Hold SHIFT + drag the left mouse button or use the arrow keys to rotate molecule(s) around the axis perpendicular (normal) to the plane of the screen.
- Hold CTRL + drag the left mouse button or use the arrow keys to move molecule(s) sideways without rotation.

Resizing Molecule Graphics

In an active Graphics Pane, you can resize regions or the entire molecule in several ways:

- Hold CTRL+ SHIFT + drag the left mouse button or use arrow keys to zoom in or zoom out.
- To enlarge a selected region or feature of the molecule, click the Fit Marked Atoms to Window button ( ). To return to the normal view, click the Fit to Window button ( ).
- Expand the Graphics Pane to fill your entire monitor screen by clicking on the Switch Screen Mode button ( ). To return to normal view, open the shortcut menu by right clicking on the graphics screen and choose Full Screen Mode, returning you to normal view or click the ESC key.
- Right-click on the Graphics Pane and choose Stereo Mode from the shortcut menu to view the graphic in stereo.

For more details about viewing graphics, see the sections below.
Sequence Pane

To activate the Sequence Pane, click on the pane or choose **Window > Sequence Pane** from the menu.

From this pane you can mark and broadcast into the Graphics Pane single or multiple selections of a protein, DNA or RNA sequence. Menu options and active toolbars in this pane are similar to those described for the other panes.

**Viewing/Manipulating 3D Graphics**

The obvious value of 3D-Mol lies in viewing and manipulating the graphics of 3-dimensional structures. Click on the **View** menu with the Graphics Pane activated to select among the viewing options described below.

**Manipulating Graphics Using Menu Commands**

**Note:** Many of the following commands are also available from toolbar buttons and the shortcut menu, launched by right clicking on the Graphics Pane.

For optimizing viewing, several options for the representation of atoms, backbone, coloring scheme, and mark types can be selected. The selections made in Atom, Backbone and Colors menus are applied to all parts of molecule, whether visible or hidden.

**View > Atoms**

These options govern the display of atoms in the molecule:

<table>
<thead>
<tr>
<th>Command</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>Represent atoms as lines</td>
</tr>
<tr>
<td>Stick</td>
<td>Represent atoms as sticks</td>
</tr>
<tr>
<td>Ball and Stick</td>
<td>Represent atoms as balls on sticks</td>
</tr>
<tr>
<td>Space Fill</td>
<td>Space-filling representation</td>
</tr>
<tr>
<td>Polyhedron</td>
<td>Represent atoms as balls without bonds</td>
</tr>
</tbody>
</table>

*Table 37.2  Molecule atom display options*

**View > Backbone**

These options control the display of the molecule backbone.

<table>
<thead>
<tr>
<th>Command</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>Switch backbone display off</td>
</tr>
<tr>
<td>CaWire</td>
<td>Represent backbone as a wire connecting Ca atoms</td>
</tr>
<tr>
<td>CaStick</td>
<td>Represent backbone as sticks connecting Ca atoms</td>
</tr>
<tr>
<td>Line Ribbon</td>
<td>Represent backbone as three-line ribbon</td>
</tr>
<tr>
<td>Flat Ribbon</td>
<td>Represent backbone as flat (thin) ribbon</td>
</tr>
<tr>
<td>Solid Ribbon</td>
<td>Represent backbone as solid (thick) ribbon</td>
</tr>
<tr>
<td>Tube</td>
<td>Represent backbone as a tube</td>
</tr>
</tbody>
</table>

*Table 37.3  Molecule backbone display options*
View > Surface
You can define the surface appearance for the structure graphics with the following options:

<table>
<thead>
<tr>
<th>Command</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td>Switch surface display off.</td>
</tr>
<tr>
<td>Points</td>
<td>Represents molecule surface as points.</td>
</tr>
<tr>
<td>Lines</td>
<td>Represents molecule surface as connected lines.</td>
</tr>
<tr>
<td>Solid</td>
<td>Represents molecule surface as a solid shape.</td>
</tr>
</tbody>
</table>

Table 37.4  Molecule surface display options

Note: View > Surface commands are enabled only if a surface dimension calculation has been performed (see Calculating Protein Surfaces, page 622).

View > Color Scheme
You can define the color scheme for the structure graphics by selecting one of the following options:

<table>
<thead>
<tr>
<th>Command</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element</td>
<td>Color by chemical element</td>
</tr>
<tr>
<td>Amino</td>
<td>Color by amino acid type</td>
</tr>
<tr>
<td>Amino Charge</td>
<td>Color by amino acid charge</td>
</tr>
<tr>
<td>Amino Hydrophobicity</td>
<td>Color by amino acid hydrophobicity</td>
</tr>
<tr>
<td>Amino Flexibility</td>
<td>Color by amino acid flexibility</td>
</tr>
<tr>
<td>Structure</td>
<td>Color by secondary structure</td>
</tr>
<tr>
<td>Temperature</td>
<td>Color by temperature</td>
</tr>
<tr>
<td>Chain</td>
<td>Color by chain (different molecules colored differently)</td>
</tr>
<tr>
<td>Clash</td>
<td>Color by clashing</td>
</tr>
<tr>
<td>Valant Angles</td>
<td>Color by valent angle</td>
</tr>
<tr>
<td>Color Schemes…</td>
<td>Create your own custom color scheme (see below).</td>
</tr>
</tbody>
</table>

Table 37.5  Graphics color scheme options

Creating Custom Color Schemes
The View > Color Scheme > Color Schemes… command allows you to created custom schemes. Once you create a custom color scheme, it appears on the View > Color Scheme menu below the stock color scheme choices.
To create a custom color scheme, launch the Color Schemes dialog box (Figure 37.7) using the above command or choose its counterpart from the Graphics Pane shortcut menu.

Previously created color schemes appear in the Schemes list on the left. Click the **Create New** button to begin creating a custom color scheme. The New Color Scheme dialog box displays (Figure 37.8):

You can choose one of three scheme types for your new color scheme, By Element, By Amino Acid or By ID. Name the color scheme, choose a scheme type and press the **OK** button. The new scheme name appears in the Color Schemes dialog box and when it is highlighted the colors for the scheme appear on the left in the Colors list (Figure 37.9):

To delete a user-defined color scheme, highlight it in the Schemes list and press the **Delete** button. To delete all user-defined color schemes, press the **Delete All** button. To save the current user-defined color scheme setup to a file, press the **Save** button. To load a previously saved Color Scheme Setup file, press the **Load** button.

Initially, default colors are assigned to the elements, amino acids or IDs in a custom color scheme. To customize the color of an element, amino acid or ID color in a custom color scheme, highlight the color scheme name in the Schemes list, select the element, amino acid or ID in the
Colors list and press the **Change** button. The Color dialog box displays. Click the **Define Custom Colors**… button to fully extend the Color dialog box (Figure 37.10):

![Color dialog box](image)

**Figure 37.10 Color dialog box**

Select a basic or custom color by clicking on its square in the Basic Colors or Custom Colors section, respectively.

**Note:** Initially, the Custom Colors palette contains only white color samples.

Define a custom color by clicking or clicking and dragging anywhere in the custom colors rainbow panel on the right. Refine the hue by dragging the slide arrow on the very right up or down. You can also define a custom color by typing the HSL or RGB values in the appropriate boxes in the custom colors section. A sample of the current custom color appears in the Color/Solid box. Press the **Add to Custom Colors** button to add the new color to the Custom Color palette on the left.

Once you have defined/selected the color, click the **OK** button to accept the color for the highlighted element/amino acid/ID in the Color Schemes dialog box. The new color appears by the name of the selected element/amino acid/ID in the Colors section of the Color Schemes dialog box.

**View > Molecules**

This option allows you to select the molecule (box checked) or hide it (box unchecked). The molecule(s) listed are those included in the displayed structure.

**View > Measure Mode**

Three measuring tools can be used in 3D-Mol to measure distances and angles in the molecule(s). These tools are accessed from either toolbar buttons or the **View > Measure Mode** menu, converting 3D-Mol to Measure Mode.

**Note:** To return to Select Mode, click the **Select Mode** button ( ) or select the option from the **View > Measure Mode** menu.

<table>
<thead>
<tr>
<th>Measure Mode</th>
<th>Icon</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off</td>
<td><img src="image" alt="Icon" /></td>
<td>Indicates Measure Mode is Off</td>
</tr>
<tr>
<td>Distance</td>
<td><img src="image" alt="Icon" /></td>
<td>Measure the distance between two atoms in Ångstroms. To measure the distance, click on the button, then click on successive atoms on the structure. The Ångstrom units appear on the screen.</td>
</tr>
</tbody>
</table>

*Table 37.6 Measuring tool options*
Note: You may find clicking on atoms easier if your image is in the ball and stick mode, rather than the wire-frame mode (select Atoms > Ball and Stick from the shortcut menu launched by right clicking on the Graphics Pane). Also, it is often helpful to mark the region to be measured, then zoom to that region before proceeding with your measurements.

**View > Mark Type**

**View > Mark Type** selects an appropriate mark tool. Click on Set Radius to define the radius within which all atoms, rotamers or monomers are to be displayed (if those options are selected) in the Graphics Pane. Structures with at least one atom within chosen radius are selected. Select each of the following options as desired:

<table>
<thead>
<tr>
<th>Command</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atom</td>
<td>Mark individual atom by clicking on it</td>
</tr>
<tr>
<td>Atoms in radius</td>
<td>Mark all atoms in radius</td>
</tr>
<tr>
<td>Monomers</td>
<td>Mark monomer (amino acid or nucleotide) by clicking on one of its atoms</td>
</tr>
<tr>
<td>Monomers in radius</td>
<td>Mark monomers in radius</td>
</tr>
<tr>
<td>Side Chain</td>
<td>Mark side chain by clicking on one of its atoms</td>
</tr>
<tr>
<td>Side Chain in radius</td>
<td>Mark side chains in radius</td>
</tr>
<tr>
<td>Molecules</td>
<td>Mark whole molecule by clicking on one of its atoms</td>
</tr>
<tr>
<td>Secondary Structure</td>
<td>Mark secondary structure by clicking on one of its atoms</td>
</tr>
<tr>
<td>Set radius</td>
<td>Define the mark radius in Å, then click OK</td>
</tr>
</tbody>
</table>

**Table 37.8 Mark display options**

<table>
<thead>
<tr>
<th>Command</th>
<th>Icon</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear/Unmark</td>
<td></td>
<td>Clear all marks</td>
</tr>
</tbody>
</table>
Table 37.8 Mark display options (continued)

<table>
<thead>
<tr>
<th>Command</th>
<th>Icon</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invert</td>
<td><img src="image" alt="Icon" /></td>
<td>Reverse the marks, meaning it unmarks current markings and marks those currently unmarked</td>
</tr>
</tbody>
</table>

View > Show

These options control atom display.

Table 37.9 Graphics Pane atom show options

<table>
<thead>
<tr>
<th>Command</th>
<th>Icon</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All atoms are shown</td>
</tr>
<tr>
<td>Marked Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All marked atoms are shown</td>
</tr>
<tr>
<td>Unmarked Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All unmarked atoms are shown</td>
</tr>
</tbody>
</table>

View > Hide

These options control atom display.

Table 37.10 Graphics Pane atom hide options

<table>
<thead>
<tr>
<th>Command</th>
<th>Icon</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All atoms are hidden, leaving the back bone displayed</td>
</tr>
<tr>
<td>Marked Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All marked atoms are hidden</td>
</tr>
<tr>
<td>Unmarked Atoms</td>
<td><img src="image" alt="Icon" /></td>
<td>All unmarked atoms are hidden</td>
</tr>
</tbody>
</table>

Graphics Pane 3D Options Dialog Box

The 3D Options dialog box allows you to set various parameters and preferences that govern the appearance of molecules displayed in the Graphics Pane. To launch the 3D Options dialog box, choose Tools > 3D Options from the menu, click the 3D Options button ( ![Icon](image) ) or choose the 3D Options command from the Graphics Pane shortcut menu. The 3D Options dialog box has five tabs that control parameters/preferences for various categories of Graphics Pane display. The Set Default button on each tab restores default parameter values for that tab. The five tabs are described below.
**Visuals Tab**

The **Visuals** tab (Figure 37.11) allows you to define parameters for graphics color/color density and font style/size.

![Visuals tab of the 3D Options dialog box](image)

**Figure 37.11 Visuals tab of the 3D Options dialog box**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Transparency</td>
<td>Determines the transparency of surface calculation solid visual representation. The higher the number, the greater the transparency.</td>
</tr>
<tr>
<td>Atoms CPK Transparency</td>
<td>Determines the transparency for atom space fill mode. The higher the number, the greater the transparency.</td>
</tr>
<tr>
<td>Fade</td>
<td>Determines color fading from foreground to background. The higher the number, the greater the fading.</td>
</tr>
<tr>
<td>Colors</td>
<td>Determines the colors for background, marks and measures. Double-click on the listing or highlight it and press the <strong>Change</strong> button to change its color.</td>
</tr>
<tr>
<td>Fonts</td>
<td>Controls the font size and type face for measure text in the Graphics Pane. Press the <strong>Measure</strong> button to change the current font style.</td>
</tr>
</tbody>
</table>

*Table 37.11 Visuals tab parameters*

**Draw Tab**

The **Draw** tab (Figure 37.12) controls atom and backbone size and width parameters.

![Draw tab of the 3D Options dialog box](image)

*Figure 37.12 Draw tab of the 3D Options dialog box*
The **Mark tab** (Figure 37.13) determines how atoms and other features are selected.

![Mark tab of the 3D Options dialog box](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ball size</td>
<td>Ball size for ball on stick representation.</td>
</tr>
<tr>
<td>Stick size</td>
<td>Stick size for stick and ball on stick representation.</td>
</tr>
<tr>
<td>CPK scale</td>
<td>Scale of space fill representation in relation to element’s atomic radius.</td>
</tr>
<tr>
<td>Line width</td>
<td>Line width. Choose a number between 1 and 5.</td>
</tr>
</tbody>
</table>

*Table 37.12 Draw tab parameters*

**Mark Tab**

The **Mark** tab (Figure 37.13) determines how atoms and other features are selected.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atom</td>
<td>Mark individual atom by clicking on it.</td>
</tr>
<tr>
<td>Atoms in radius</td>
<td>Mark all atoms in radius.</td>
</tr>
<tr>
<td>Molecule</td>
<td>Mark whole molecule by clicking on one of its atoms.</td>
</tr>
<tr>
<td>Monomer</td>
<td>Mark monomer (amino acid or nucleotide) by clicking on one of its atoms.</td>
</tr>
<tr>
<td>Monomer in radius</td>
<td>Mark monomers in radius.</td>
</tr>
<tr>
<td>Side chain</td>
<td>Mark sidechain by clicking on one of its atoms.</td>
</tr>
<tr>
<td>Side chain in radius</td>
<td>Mark sidechains in radius.</td>
</tr>
<tr>
<td>Secondary</td>
<td>Mark secondary structure by clicking on one of its atoms.</td>
</tr>
<tr>
<td>Mark radius</td>
<td>Set selection radius in Angstroms.</td>
</tr>
</tbody>
</table>

*Table 37.13 Mark tab parameters*

**Note:** The options on the Selection Tab duplicate options available as submenu items under View.
**Conflicts Tab**

The **Conflicts** tab (Figure 37.14) parameters define clash and valent angle interference display in the Graphics Pane.

![Conflicts tab of the 3D Options dialog box](image)

**Parameter** | **Function**
--- | ---
Clash distance | Sets minimum distance for clashing.
No clash distance | Sets maximum distance for clashing.
Hydrogen atoms can clash | Check to allow hydrogen atom clashing.
Non valid valent angle | Sets minimum for an angle to be considered a valent angle.
Valid valent angle | Sets maximum for an angle to be considered a valent angle.

*Table 37.14 Conflicts tab parameters*

**Actions Tab**

The **Actions** tab (Figure 37.15) controls general interface behavior in the Graphics Pane.

![Actions tab of the 3D Options dialog box](image)
Stereo Viewing Option

3D-Mol provides a stereo viewing option in the Graphics Pane. Right-click in the Graphics Pane and choose **Stereo Mode** from the shortcut menu to toggle in and out of Stereo Mode.

Sequence Pane Viewing Options

The Sequence Pane Viewing Options dialog box (Figure 37.16) allows you to define the appearance of the sequence in the Sequence Pane. To launch the Sequence Pane Viewing Options dialog box, right-click on the Sequence Pane and choose **Options** from the Sequence Pane shortcut menu.

![Sequence View Options dialog box](image)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Block</td>
<td>Number of residues that are grouped together.</td>
</tr>
<tr>
<td>Blocks per Line</td>
<td>Number of blocks that appear on one line in the Sequence Pane.</td>
</tr>
<tr>
<td>Set Default</td>
<td>Resets sequence view options to 3D-Mol default values.</td>
</tr>
</tbody>
</table>

Table 37.16  Sequence View Options parameters

**3D-Mol Tools**

3D-Mol includes two tools, one for aligning molecules and one for calculating the surface dimensions of a molecule. Using the alignment tool, you can align any two chains of ten monomers or greater in the current 3D-Mol project. You can calculate and display surface dimensions of a molecule via two different methods, the Connolly or Varshney methods.

Aligning Chains

3D-Mol includes an alignment tool for aligning any two protein chains of greater than ten monomers in the project.
Performing an Alignment

To align two chains, select one chain (that is greater than ten monomers) in the Structure Pane and choose Tools > Align Chain from the menu.

Note: If you select a chain that is ten monomers or shorter for alignment, an error message is generated.

The Select Chain to Align dialog box (Figure 37.17) displays allowing you to choose a second chain from the list of chains currently in the project.

Viewing Alignment Results

When the alignment process is finished, the results appear in the 3D-Mol window (Figure 37.18):

The aligned molecules are shown side-by-side in the graphics pane with their names above. Notice that an Alignments folder has been added to the Project Pane and the new alignment, NewAlign<#>, is listed within. The sequence for each of the aligned chains is listed on its own tab in the Sequence Pane.

There are two modes in which you can view alignments, Align Backbone and Align Atoms modes. By default the alignment initially appears in Align Atoms mode, which shows the backbones and associated atoms for the two aligned molecules side-by-side (Figure 37.18). You can switch to Align Backbone mode by pressing the View Aligned Backbone button ( ) or by
choosing Align Backbone Mode from the Graphics Pane shortcut menu. Align Backbone mode shows only the backbones of the aligned molecules superimposed on each other (Figure 37.19).

**Figure 37.19 Align Backbone mode**

Switch back to Align Atoms mode by pressing the View Aligned Atoms button ( ) or by choosing Align Atoms Mode from the Graphics Pane shortcut menu.

**Working with Alignments**

Alignments are considered a type of structure in 3D-Mol and can be manipulated, exported and saved to files and to the project in the same way as imported structures and NewEntries (see the appropriate sections in this chapter).

**Calculating Protein Surfaces**

Protein surface topography analysis is an important part of structure-based drug design, because drugs tend to bind in crevices on protein surfaces. An important first step in analyzing protein topography is defining the surface for a protein molecule. 3D-Mol includes a tool for calculating and displaying a graphical depiction of the surface of a molecule. 3D-Mol provides two different methods of surface calculation.

To calculate the surface for the chain(s) in a structure, display the structure in the 3D-Mol workspace and choose Tools > Calculate Surface from the menu. The Surface Calculation dialog box displays (Figure 37.20):

**Figure 37.20 Surface Calculation dialog box**

Select the surface calculation method, Connolly or Varshney (for references, see Appendix B), and designate a probe radius between 1 and 10 angstroms (probe radius determines the dis-
distance from the backbone at which the surface is calculated). Click the OK button to initiate the surface calculation.

**Notes:**

- Generally, the larger the molecule, the longer the surface calculation time. Large molecules may require several minutes for the calculation to complete.
- If you want to cancel the surface calculation once it has been initiated, choose **Tools > Cancel Surface Calculation** from the menu.

When the calculation is finished, the calculated surface displays on the molecule in the Graphics Pane in Line view mode (Figure 37.21).

![Image](image.png)

**Figure 37.21 Varshney surface in Line view mode**

Choose **View > Surface > <surface view command>** from the menu to change or remove the surface view (see **View > Surface**, page 612). You can modify surface viewing characteristics on the **Visuals** tab of the 3D Options dialog box (see **Graphics Pane 3D Options Dialog Box**, page 616).

To recalculate a surface, choose **Tools > Recalculate Surface** from the menu.

**Note:** As indicated above, the surface calculation menu item adopts one of three forms, depending on the stage of surface calculation. Before a surface has been calculated, the Tools menu command appears as **Calculate Surface**. During surface calculation, the menu command appears as **Cancel Surface Calculation**. After a surface has been calculated, the menu command appears as **Recalculate Surface**.

### Exporting and Saving 3D-Mol Data

3D-Mol structures can be exported to files or opened in Vector NTI and saved to the Vector NTI Database. 3D-Mol projects can be saved as 3D-Mol project files (.p3d).

### Opening/Importing Structures from 3D-Mol in Vector NTI

To open or import a structure into Vector NTI, highlight the structure in the Project Pane and choose **File > Send to Vector NTI** from the menu. This launches Vector NTI and displays an Open/Import dialog box showing a list of the chains in the selected structure. From the Open/Import dialog box, you can choose which chain(s) from the structure you want and whether you want to open in Vector NTI or import directly into the Vector NTI Database. Highlight the desired chain(s) in the Open/Import dialog box and press the appropriate button, **Open** or **Import** to complete the action. Vector NTI opens or imports each selected chain as a separate molecule. If you choose to open the chain(s) in Vector NTI, you can subsequently save the chain(s) to the Vector NTI Database from their Molecule Viewing windows.
Exporting Structures to a File

To save a 3-D Structure to an independent file, highlight the structure in the Project Pane and press the Save button or select File > Export Entry from the menu. You can save the structure as a PDB or M3D file.

Saving 3D-Mol Sessions

To save a 3D-Mol session, press the Save button or select File > Save Project or Save Project As. The session is saved as a .p3d file.

Note: Upon saving the project, if you have modified NewEntries and have not yet saved them, 3D-Mol asks you to save them first before it saves the project.
Introduction

This chapter describes Vector NTI’s Tools, powerful vehicles for extending Vector NTI’s functionality and providing connections to external programs and Internet services. The chapter covers only general configuration topics. Each particular tool has its own description in other chapters in this user’s manual.

Vector NTI Advance is shipped with more than 20 tools providing connection to various Internet servers. You can add your own tools connecting Vector NTI to local or remote programs using the configuration mechanism described in this chapter.

Configuring Vector NTI Tools

All registered tools can be accessed in the main menu options appropriate to the tools such as Analyses, Align and Tools. In addition, Vector NTI stores the information describing currently registered tools in the Tools Manager, where you can also create new tools or links to tools and perform operations on existing tools.
To open Tools Manager, select **Tools > Tools Manager** in any Vector NTI Advance application, including Database Explorer.

![Figure 38.1 Tools Manager](image)

The Tools Manager (Figure 38.1) displays folders containing all Vector NTI tools. Click on a folder to open it, where its respective tools are displayed. Each folder and its tool contents correlate with corresponding menu options on the Vector NTI Advance toolbars. The Info pane in this dialog box displays, for a selected tool, its name, function, and input type.

Buttons in the dialog box perform the following operations:

- **Create Tool** opens the Tools Parameters dialog box with two modes:
  - To create a new tool from scratch, press the **Create Tool** button with a folder selected. The dialog box is empty except for the folder name from which it was launched.
  - To create a new tool based on an existing tool, select the tool in the left pane and press the **Create Tool** button. In this instance, the dialog box opens loaded with data for that tool. All fields can be edited.
- **Rename** enables the text box of a selected tool where you can rename it
- **Create Folder** creates a new folder in the list pane with an enabled text box for name entry
- **Delete** deletes a selected folder or tool

In addition, tools with a .tq extension can be edited. To modify a tool or its description, double-click on a selected tool. This, too, opens the Tools Parameters dialog box (Figure 38.2). *The name text box is disabled if opened in this manner.*

![Figure 38.2 Tools Parameters dialog box](image)
Vector NTI displays the Tools dialog box with the following sections:

The Tool Parameters dialog box consists of five fields:

- **Name** Use the vertical bar character | to give your command a “segmented” name to arrange a nested menu structure. The first part of the name (from the first character to the first vertical bar) becomes the name of the menu item in the Tools menu, the second part, from the first to the second vertical bar, becomes the name of the submenu item in the first submenu, etc. The last segment of the name (from the last vertical bar to the last character of the name) becomes the actual name of the command item, associated with the tool description.

- For logical access to analysis tools, Vector NTI Advance places all tools with names starting with “Align | ” at the end of the Align menu, etc.

- **Description** Contains a one-line description you can edit. This line is displayed on the status bar when you browse through the tools menus. You can save the edited description under a new or existing name. If saved under an existing name, the tool description is replaced by the new description.

- **Input Type**: Lets you select the type and contents of the input information file Vector NTI prepares for the tool before running it. The most frequently used Input Type options are described below:

<table>
<thead>
<tr>
<th>Input Type</th>
<th>Input information</th>
</tr>
</thead>
<tbody>
<tr>
<td>(none)</td>
<td>The tool requires no input</td>
</tr>
<tr>
<td>&lt;N&gt; &lt;Type&gt; Molecule(s) in &lt;Format&gt; format</td>
<td></td>
</tr>
<tr>
<td>&lt;N&gt; is “Active”, “Two”, “One or more”, or “Two or more”</td>
<td></td>
</tr>
<tr>
<td>&lt;Type&gt; is “DNA/RNA” or “Protein”</td>
<td></td>
</tr>
<tr>
<td>&lt;Format&gt; is “GenBank/GenPept” or “EMBL/SWISS-PROT”</td>
<td></td>
</tr>
<tr>
<td>Molecules are saved into the _molembl file (EMBL/SWISS-PROT formats) or _molgbk file (GenBank/GenPept formats). If more than one molecule is selected, the file would contain multiple records. Vector NTI can prepare input from selection in the Database Explorer or from the active Display window (if just one molecule is required)</td>
<td></td>
</tr>
<tr>
<td>&lt;N&gt; &lt;Type&gt; Sequence(s) in FASTA format</td>
<td></td>
</tr>
<tr>
<td>&lt;N&gt; is “Active”, “Two”, “One or more”, or “Two or more”</td>
<td></td>
</tr>
<tr>
<td>&lt;Type&gt; is “DNA/RNA” or “Protein”</td>
<td></td>
</tr>
<tr>
<td>Sequences are saved in the _fasta file. If more than one molecule is selected, the file would contain multiple sequences. Vector NTI can prepare input from selection in the Database Explorer or from the active Display window (if just one sequence is required)</td>
<td></td>
</tr>
<tr>
<td>Active Document</td>
<td></td>
</tr>
<tr>
<td>Active Molecule in Extended GenBank Format</td>
<td></td>
</tr>
<tr>
<td>Active molecule is saved in _molgbk file (GenBank with Vector NTI extensions)</td>
<td></td>
</tr>
<tr>
<td>Link Data</td>
<td></td>
</tr>
<tr>
<td>Command string is taken from the corresponding User Field</td>
<td></td>
</tr>
</tbody>
</table>

*Table 38.1 Input Type options*

**Note:** On Windows platform, _seq, _fasta, _molembl, and _molgbk files have .txt extension. Files are created in the Tools directory/folder.

1. **Script** Allows you to select a Vector NTI script or executable program to be opened after Vector NTI finishes preparing input information according to the Input Type specification. Press the **Browse** button to select a script or program to be opened.

2. **Command**: Contains an extra parameter string given to and interpreted by the script or program you selected.
Note: Before sending a command string to a program or script, Vector NTI scans it for symbols %1…%6 and performs the following substitutions:

- %1: full path to the data file (see the table above)
- %2: name of the molecule
- %3: selection start (if selection in the Display window was used)
- %4: selection end (if selection in the Display window was used)
- %5: current author’s e-mail address from the Author Info dialog
- %6: “aa” for protein data, “na” for DNA/RNA data

When you finish entering or editing the tool fields, click OK.

Sorting Tools in Tools Manager

Items in the Tools Manager and all of its submenus are listed alphabetically, but you can sort items by clicking tools and dragging them into new positions on the list. You can even move tools to new folders. After you have made sorting changes, the corresponding menus in Vector NTI applications will reflect the changes.

Adding a Web Link to the Tools List

To add Web links to your list of tools, in the Tools Manager select the folder where the new item is to be stored. Select Open from its corresponding shortcut menu. This opens a Windows Explorer window displaying large icons for all of the tools in that folder. Open your Web browser to the page you would like the link to open. On the web browser screen, click and drag from a “Create a Web Link” icon (In Netscape 4.7, the symbol ( ) to the left of the Location text box, or in Internet Explorer, the symbol ( ) to the left of the URL in the location text box.) into the <Folder name > Explorer window.

Note: Adding or deleting command descriptions does not affect script or program files on disk, only the description of the way Vector NTI opens the script or program is affected.

Program- and Script-based Tools

Each Vector NTI Tool can be configured either to invoke a certain program or execute a script. Program-based tools are useful to open another molecular biology program to perform an extra analysis of your sequence, alignment of several sequences or some other operation, not supported by Vector NTI. These tools are platform- and environment-dependant and cannot be easily moved from one Vector NTI installation to another. Vector NTI script-based tools, on the other hand, are designed to allow the creation of portable tools easily transferable from one installation and platform to another.

Script-based tools are executed by a separate script execution module (SXM) that hides the platform-specific details from the script-writer. In addition, script execution modules serve as intermediate communication and data exchange layers between Vector NTI and such services as electronic mail or Internet servers. The basic set of high-level primitives covering data exchange between scripts and Vector NTI, electronic mail and Internet browsers is supplied together with SXM as a SXM runtime file (SXMRUN). The runtime layer allows the configuration of some functions like access to the Web browser by supporting the configuration function available as RECONFIG script and the corresponding Configure Scripts tool. This tool displays a dialog box allowing you to choose a service you want to configure and to select one of the available platform-specific Service Support Modules. For example, you can choose one of several Web Browser Interface modules matching the Web browser software installed on your computer.

Invitrogen supplies a basic set of tools/scripts with the standard distribution of Vector NTI and provides on-line access to extra tools and scripts via the Invitrogen web site (www.invitro-
gen.com/bioinformatics). The Tools page contains the latest versions and extensions of platform-dependent part of Tools support like SXM runtime files for Windows and Macintosh platforms, platform-specific service support modules etc. The Scripts page contains platform-independent scripts in a form of self-installing packages as well as technical information concerning the scripts.
Introduction

This chapter contains several of Vector NTI’s miscellaneous features.

The first three described features work well in tandem and provide a practical and efficient way to search public databases for molecules or citations, retrieve them, complete with molecule files or abstracts that can be viewed and stored. Some citations are included with Vector NTI Advance, but you can expand and store and share references and abstracts as you would any other database object.

PubMed/Entrez Search

This feature of Vector NTI Advance is a client program for the NCBI Entrez/PubMed search engines. Database searches are performed in much the same way they are in the NCBI website: through the PubMed/Entrez Search Tool, you send a user query to Entrez/PubMed. The convenience of this Vector NTI Advance tool lies in its connectivity to the Internet and its ability to retrieve and list the results. From this list, you can easily view the results in either a Molecule Display window or the Citation Viewer and store them in the Vector NTI Suite database.

Results of the query, however, are only molecule or citation summaries, not the documents themselves. Any operation that requires the entire document such as Export, Copy, Drag and Drop, or Open will require that the selected documents be downloaded first.

Note: For this feature to work in your computer, an internet connection must be available and properly configured and open. This especially applies to corporate networks where the local networks have access to the Internet only via a corporate proxy server. Although the program tries to determine whether or not to use a proxy, and the name of the proxy, the Edit > Setting > Internet Settings should be the first place to look if the program is unable to connect to the NCBI server or to download anything.
To initiate a search, choose **Start > Programs > Invitrogen > Vector NTI Suite 10 > PubMed/Entrez Search**. The PubMed/Entrez Search dialog box that opens is untitled and remains so until you save the results under a name (Figure 39.1).

![Figure 39.1 PubMed/Entrez Search dialog box](image)

Select the database to be searched in the Database drop-down menu.

The section below the drop-down menu is the Query Pane where you enter your query. It is possible to assign each search term to a database field, such as article title, authors, MeSH terms.

Click the cursor in the blank row of the term text field, enabling the search fields drop-down list at the right end of the row. Select a field to be searched. Click in the blank portion of the row and enter the search keywords, name(s), or topics.

**Notes:**

- When submitting a query like "cat"[All fields], the NCBI server translates it as "tomography, x-ray computed"[MeSH Terms] OR "cats"[MeSH Terms] OR cat[Text Word] and searches for documents matching this translated query. If a space is to be entered in your text string, be sure and enclose the entire string in quotation marks.
- If, when you are entering or editing text, you press ENTER by mistake, the query will be submitted. CTRL + ENTER also causes the query to be submitted.

To add a row to enter a new search condition, select **Insert Row** from the shortcut menu, select **Edit > Insert Row**, or press the space bar, and enter new text. In the drop-down menu to the left of the row, select **Or**, **And** or **Not** to qualify the search conditions.

**Defining Search Conditions**

Before initiating the search, select **Edit > Settings**, opening the Settings dialog (Figure 39.2).

![Figure 39.2 Settings dialog box](image)
In the **Common** section, specify the number of document summaries acceptable per results page and define the History buffer. Check the **Save Results**… box to save downloaded summaries when you save a query.

The **History** buffer contains the last queries and query results. You can move through the history by using **Back** and **Forward** buttons on the PubMed/Entrez Search dialog box.

In the **Publication Date** section, enter publication dates to narrow the search to a certain date range. In the **Entrez date** section, enter acceptable molecule submission dates.

Click **OK** to return to the PubMed/Entrez Search dialog box.

Select **Edit > Internet Settings** to open the Internet Connection dialog box (Figure 39.3).

![Figure 39.3 Internet Connection dialog box](image)

**Internet Connection** options allow you to specify whether or not connection to the Internet should be performed via a proxy server. If you select the Use Internet Explorer settings radio button, Vector NTI will attempt to detect and use your default settings to connect to the Internet. If the default settings detection is not successful, select either **Direct connection** if you do not use a proxy server or **Use proxy server** if you use a proxy server. In the case of a proxy server, enter the proxy name or address and port number. If the proxy requires a password, check the **Proxy requires authentication** box and enter the user name and password. *This facilitates access for the search through a proxy server.*

Click **OK** to return to the PubMed/Entrez Search dialog box.

When you have entered all search conditions, click the **Search** button ( ). The search results are displayed in the list pane in the bottom half of the dialog box.

**Reviewing Search Results**

Article or molecule summaries are listed in the Results Pane. There are several techniques to download the citation or molecule results:

- Double-click on its identification code for the selected object or select multiple objects
- Select **Open <Article/Molecule> ID <#>** from the shortcut menu associated with the object. DNA and protein database search results open in a Vector NTI Display window. A citation abstract opens in the Citation Viewer.
- Choose **Copy** from the shortcut menu. The item is downloaded and placed on the Clipboard for pasting into Vector NTI Database Explorer. The object is downloaded.
- Select **Export to File** from the shortcut menu to download the selected item and places it in a file you specify.
- Drag and drop selected object(s) from the Results Pane into Database Explorer or Windows Explorer folders.

To open multiple search objects, select them from the search results list, and, depending on the object type, drag and drop them into the Database Explorer, or the Citation Viewer windows. Multiple objects open in cascaded multiple instances of the display windows.
If the search was performed against the PubMed database, the shortcut menu associated with a selected search result lists a **Related Articles** option. Select this command to download abstracts related to the selected article. The Query Pane changes to read **Related Articles for Article ID <#>**. In this case, you cannot modify the Query Pane. Actions in the Results Pane are not affected.

**Saving Search Results**

If you select the **Save** option from the Search dialog box, the only thing that is saved is the set of search parameters for launching future searches, with a .ps6 extension.

Select the object(s), choose **Save Parameters As** and in the dialog box that opens, name the file and its destination. You may want to save results in the Citation folder under Vector NTI program files. To save the molecule or document abstract, open it first in the Molecule Display window or in Citation Viewer and save your object from that location. To open the citation in Citation Viewer, refer to details in the Citation Viewer section, page 634.

**Modifying the PubMed/Entrez Display Window**

To modify the column display in the results list in this window, open the shortcut menu with the cursor on the column headings and select **Columns**. In the Columns dialog box that opens, using the forward (>>) and reverse (<<) arrows, move column topics to the Selected Columns list for display, or move them back to the Available columns list to remove them from display. The descending order in the **Selected Columns** list corresponds to left to right display in the results pane of the dialog box.

Items listed in the Results Pane columns can be sorted by clicking on a column heading. Click again for reversing the list.

Right-click to the right of all of the buttons on the Rebar (toolbar) to open a popup menu where you can turn on/off text labels and the toolbar itself.

**Citation Viewer**

This feature of Vector NTI Advance provides a vehicle for displaying and reviewing citation abstracts downloaded by PubMed Search or created manually and stored in the Vector NTI database or local file. Citations can be printed, edited and saved from Citation Viewer.

To open Citation Viewer, select Citation Viewer in the program group or folder where you installed Vector NTI. If opened from Database Explorer, Citation Viewer is empty and must be loaded from other sources, listed below:

From the PubMed/Entrez Search results, or from the Citations Table in Database Explorer, drag and drop citations into Citation Viewer. *It helps if all of your screens are in “Partial Screen Size” to do this.* If you transfer multiple files, all of the abstracts will open in Citation Viewer. You can alternate between references by clicking the **Back** and **Forward** buttons on the toolbar (they resemble the Back and Forward buttons on your Web browser) or you can select from the list of open citations under the Windows menu option.

**Opening Citation Viewer**

From the Start menu, choose **Start > Programs > Invitrogen > Vector NTI Suite 10 > Citation Viewer**. Alternatively, you can open an individual citation into the Citation Viewer window (Figure 39.4) from the PubMed/Entrez Search display window or from the Citations Table in Database Explorer. Simply double-click the entry in either location to load Citation Viewer.
Citation viewer registers its file types, so double-clicking on a .ca6, .cq or .ct file in Windows Explorer will cause Citation Viewer to open displaying that file.

Figure 39.4 Citation Viewer

Elements of Citation Viewer

The Citation Viewer display window shares many of the same features found in other Vector NTI Advance applications:

- The window title bar displays the name of the open object.
- Click the maximize box ( ) in the upper right corner on the title bar to maximize the window, filling the entire Vector NTI workspace.
- A display window is divided into two panes: a Text Pane containing folders with descriptions of the citation such as author, source, and public database identification data. The Abstract Pane displays the abstract for the citation.
- A vertical, movable split bar divides the two panes.
- All Vector NTI Advance windows have a Main Menu and toolbars where many different options and operations can be initiated. Most of the toolbar buttons in Citation Viewer are similar to those you have seen in other display windows. Most toolbar buttons have associated tool tips that explain the button function. Hover the cursor over the button to display the tool tip for that button.

Back and forward buttons will bring up the previous or next viewed document if it is still open.

Note: If the citation contains an URL to its full text or abstract in the Internet, a link to that URL will be displayed in Text Pane. Clicking on it causes the default Internet browser to open to the appointed location.

The following feature can be performed in Citation Viewer. Refer to Chapter 3 for details:

- Copying selected data to the clipboard
- Exporting data to external programs
- Printing data
- Modifying the Citation Viewer layout
- Saving data
Remote Database Sources Wizard

The Remote Database Sources Wizard steps you through the process of creating links to remote servers, which enable Vector NTI remote services tools to connect to external databases, such as Entrez and LabShare. Basic links for BLAST and PubMed/Entrez Searches are already created; you do not have to create them. You must create the link to your LabShare Database, however. The Remote Database Sources Wizard assists you in doing so.

Configuring the Internet Connection

Before you can create a link to your LabShare Database, you must configure your Internet connection appropriately so that a proper connection can be made between Vector NTI and your LabShare Database.

To set your Internet connection parameters and configure a path to your LabShare Database, first choose Start > Programs > Invitrogen > Vector NTI Suite 10 > Remote Database Sources Wizard to open the Data Sources (Vector NTI) folder (Figure 39.5).

![Figure 39.5 Data Sources (Vector NTI) folder](image)

To launch the Internet Connection dialog box, choose Tools > Internet Settings… from the menu in the Data Sources (Vector NTI) window (Figure 39.6):

![Figure 39.6 Internet Connection dialog box](image)

- **Use Internet Explorer Settings** – If a proxy server is not used for accessing your LabShare Database server, use this setting.
- **Direct Connection** – If a proxy server is not used for accessing your LabShare Database server and the Use Internet Explorer Settings selection does not configure a connection to your LabShare Database server, use this setting.
- **Use Proxy Server** – If a proxy server is used to access your LabShare Database server, select Use Proxy Server and enter the appropriate information.

After configuring your Internet Connections appropriately, close the Internet Connection dialog box to return to the Data Sources (Vector NTI) folder.

**Note:** You need only configure your Internet Connection settings once, as long as your Internet connection configuration does not change.
Creating Remote Database Links

**Note:** Before you create a remote link to your LabShare Database, you must configure your Internet connection, if you have not already done so (see above).

To configure a path to your LabShare Database, in the Data Sources (Vector NTI) folder window, double-click the **Add Remote Data Source** icon to launch the Remote Data Sources Wizard (Figure 39.7):

![Figure 39.7 Remote Database Sources Wizard, screen 1]

Select **LabShare Server at URL** and type in the path to that server.

**Important:** The path should EXCLUDE the name of the LabShare cgi script. For example, http://myServer.myDomain/cgi-bin/ve/ is a correct path designation, whereas http://myServer.myDomain/cgi-bin/ve/ve-cgi.cgi is not. If you include the name of the cgi script in the path designation, Vector NTI will not be able to connect to the remote database.

If your LabShare server requires a password, check **LabShare Server Requires Authentication** and complete the server authentication section. Click the **Next >** button to continue.

Select the Data Type(s) for this path configuration (Figure 39.8):

![Figure 39.8 Remote Database Sources Wizard, screen 2]

**Important:** For the LabShare remote database, all Data Types (Database Tables) are included in a single link configuration. You DO NOT have to configure a separate link for each Data Type, so the choices are pre-selected.

Click the **Test** button to launch the Test URL dialog box. This tests the connection to the LabShare Database and ensures that the path setting has been configured correctly.
If the connection is successful, objects are downloaded to the Object List Pane in the Test URL dialog box. If the connection is not successful, an error message, such as the following is generated (Figure 39.9):

![Image of an error message](vector_nti_advance_10_user_manual.png)

*Figure 39.9 Internet connection error*

If you receive an error message upon testing, try one or both of the following:

- Step back through the Remote Database Source Wizard and reconfigure the path (see Figure 39.7 above).
- Launch the Internet Connection dialog box and reconfigure your Internet connection settings (see Configuring the Internet Connection above).

Click the **Next >** button to continue. In the next dialog box (Figure 39.10), choose the protocol options that your server supports.

**Note:** All protocol options are supported for LabShare servers, so all choices are pre-selected.

![Image of the Remote Database Sources Wizard](vector_nti_advance_10_user_manual.png)

*Figure 39.10 Remote Database Sources Wizard, screen 3*

- **RETRIEVE** – Enables the ability to download objects from the LabShare Database to the Mini-Exploring window (see Accessing Remote Databases in Vector NTI).
- **SAVE** – Enables the ability to save objects to the LabShare Database via the Mini-Exploring window (see Accessing Remote Databases in Vector NTI below).
- **Support for Batch Downloads** – Enables batch-downloading capability from a LabShare Database into Vector NTI from the Mini-Exploring window (see Accessing Remote Databases in Vector NTI below).
- **Support for VNTI Archives** – Enables the ability to include links to the parent molecules of constructed/designed molecules when transferring constructed/designed molecules to the LabShare Database.

**Note:** Links to parent molecules are included, not the actual parent molecules.
Click the **Next >** button. Type in an appropriate name for the Data Source (Figure 39.11) and click the **Finish** button to complete the configuration.

![Figure 39.11 Remote Database Sources Wizard, screen 4](image)

A new icon for the newly configured Data Source is added to the Data Sources (Vector NTI) folder.

Close the Data Sources (Vector NTI) folder.

**Accessing LabShare Databases in Vector NTI**

LabShare Databases can be accessed directly in Vector NTI using the LabShare Mini-Exploring window. You can launch the LabShare Mini-Exploring window from the Vector NTI workspace window by choosing **File > Open** and selecting the **Remote Sources** tab of the Open dialog box (Figure 39.12).

**Note:** Mini-Exploring window capability is installed in Vector NTI Advance 10 as part of a Typical installation. If you do not have a LabShare Database and, therefore, do not want/need Mini-Exploring window capability, when installing Vector NTI, choose the **Custom** installation option and uncheck the **Remote Access Support** checkbox.

![Figure 39.12 LabShare Mini-Exploring window](image)

**Navigating the Mini-Exploring Window**

Initially, the Data Source box shows the LabShare Database Table you have most recently accessed (if any) and the Filter (Subset) box is blank; objects are not yet downloaded to the Object List Pane in the Open dialog box.

There are three ways to download objects from the LabShare Database into the Object List Pane:
- Click the Go button ( ) to download the molecules in the Main base of the Table that is currently shown in the Data Source box.
- Choose the Database Table you wish to view from the Data Source drop-down box list (and click the Go button if objects do not automatically load in).
- To download the objects in a particular subset of the Table currently shown in the Data Source box, select that subset from the Filter drop-down box (see the following details regarding subsets shown in the Filter drop-down list).

When the Open dialog box is first opened (before objects are downloaded into the Object List Pane), the Filter drop-down list initially shows the list of subsets you have most recently viewed (History List), in the order of most recently viewed beginning at the top of the list. The icon next to each subset indicates that you are viewing the History List and not the actual LabShare subset listing (Figure 39.13):

![Figure 39.13 Viewing the History List in the Mini-Exploring window](image)

**Note:** Because the History List is merely a record of the subsets that have been viewed, it may list subsets that no longer exist in the LabShare Database (subsets that have been deleted since they were last viewed, for example). If such a subset is selected, a corresponding error message is generated.

Once objects are downloaded to the Object List Pane, the Filter drop-down list shows the actual list of subsets for the LabShare Database Table you are viewing, indicated by the icon next to each subset (Figure 39.14):

![Figure 39.14 Viewing a LabShare subset](image)
**Formatting the Page Size in the Mini-Exploring Window**

Objects are downloaded to the Mini-Exploring window one page at a time; the default page size is 100 objects. You can download the next or previous page of objects by pressing CTRL + PAGE DOWN or CTRL + PAGE UP, respectively, on your keyboard or by clicking the DOWN ARROW on the right side of the ( ) button and choosing the commands from the menu (Figure 39.15).

![Figure 39.15 Mini-Exploring window paging menu](image)

Use the Go To dialog box, launched by choosing Go To from the menu or by clicking the Go To ( ) button, to download a specific page of objects or to change the number of objects displayed per page. Enter the specified information and click OK.

**LabShare Subset and Object Selection Options**

After you have downloaded data to the Mini-Exploring window, the paging menu changes to include two more commands (Figure 39.16):

![Figure 39.16 Subset and Options commands on paging menu](image)

- **Subset** – Allows creation of new, empty subsets.
- **Options** – Launches the Object Check-out Options dialog box where you can set object check-out/check-in parameters (Figure 39.17):

![Figure 39.17 Object Check-out Options dialog box](image)

**Important:**

- If you do NOT check the **Check out object automatically when opened from the remote database** option, you will NOT be able to save the object back to the LabShare database.
- If you do NOT check the **Check in object automatically when saved in the remote database** option, the object will NOT be checked in when you save it back to the Lab-
Share database. In this case you must check the object in manually in the LabShare Database window (see the LabShare User’s Manual).

**Opening Objects from the LabShare Database in Vector NTI**

Once the data downloads to the Mini-Explorer window, highlight the desired object (or objects, if you have configured your LabShare path to support batch downloading) and click the OK button to open the object(s) in the corresponding Vector NTI application (i.e. DNA/RNA molecules open in Vector NTI; Citations open in Citation Viewer, etc.).

**Important:** If you want to manipulate an object (analyze, edit, etc.) that you have opened from the Vector NTI LabShare Database in a Vector NTI Advance application window, you must first save the object to your local Vector NTI Database.

**Saving Objects to the LabShare Database from Vector NTI**

Objects opened in Vector NTI applications can be saved to the LabShare Database directly in the Mini-Exploring window.

To save an opened object, perform the following steps:

1. Choose `<data type> > Save As` from the drop-down menu in the VNTI application (for example, choose File > Save As when saving a DNA/RNA molecule or Citation > Save As when saving a citation) and select the Remote Sources tab in the Save As dialog box.

2. Select the LabShare Database Table (Data Source) and/or the desired subset (Filter), if any, to which you want to save the object.

**Note:** As with the Open dialog box described above, initially the Filter drop-down list in the Save As dialog box shows a History List of most recently viewed subsets, as evidenced by the subset icons. Once a subset has been downloaded into the object list window, the Filter drop-down list shows the list of subsets in the LabShare Database, indicated by the subset icons.

Once you have selected the appropriate Table (Data Source) and subset (Filter), if any, the objects in that Table/subset display in the Object List Pane (Figure 39.18):

![Figure 39.18 Object List for selected subset](image)

1. Designate a name for the object you are saving.

2. Check **cc to Local DB** if you want to save the object to both the LabShare and VNTI Local Databases.
Note: The cc to Local DB check box is disabled for molecules opened from the Vector NTI Local Database. However, if you have modified a VNTI Local Database molecule, it is saved to the Local Vector NTI Database automatically when you save to the LabShare Database.

3. Click the OK button to complete the operation.

GCG Converter

GCG Converter is a tool in Vector NTI Advance that converts sequences in GCG file formats so that they can be imported into the software.

To open this tool, select **Start > Programs > Invitrogen > Vector NTI Suite 10 > Utilities > GCG Converter**. This opens the GCG to Vector NTI Converter dialog box (Figure 39.19) for importing and converting the GCG files.

![Figure 39.19 GCG to Vector NTI Converter dialog box](image)

Click the **Add** button, opening a standard Open dialog box where you can browse for your GCG files (Figure 39.20). Select those you want to convert and import into Vector NTI; click **Open**.

![Figure 39.20 Open dialog box where you can browse for your GCG files](image)
The files are listed in the GCG Converter list box (Figure 39.21). To convert all of the files to VNTI format, click **Convert All** or select those you wish to convert and click the **Convert Selected** button.

After the conversion process concludes, symbols appear to the left of each file name indicating the resolution of the conversion process. Molecule types or Errors are also listed in the Type column.

Shortcut menu options for each file listed are dependent upon the file symbol. For instance, you can select **Get Error Information** from the shortcut menu from each of the files showing an error message. You can also choose to remove the file from the list.

Symbol interpretations and the user solution are listed below.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Interpretation</th>
<th>User Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>Conversion was successful</td>
<td>N/A</td>
</tr>
<tr>
<td><img src="image" alt="Symbol" /></td>
<td>Sequence type could not be discerned</td>
<td>In sc menu, select <strong>Set Type</strong>. In the dialog box, in drop-down menu select DNA or Protein and click <strong>OK</strong>.</td>
</tr>
<tr>
<td><img src="image" alt="Symbol" /></td>
<td>Source file indicates one sequence type, but a wrong symbol is found in the sequence; position is specified</td>
<td>Select <strong>Open</strong> in Text Editor (Notepad or other you have specified. See Figure 39.22.). Verify and edit the error. Re-convert.</td>
</tr>
<tr>
<td><img src="image" alt="Symbol" /></td>
<td>Source file is empty or an error was detected in reading the first line in the file</td>
<td>Select <strong>Open</strong> in Text Editor (Notepad or other you have specified). Correct source file problem.</td>
</tr>
<tr>
<td><img src="image" alt="Symbol" /></td>
<td>Sequence cannot be detected in the source file</td>
<td>Select <strong>Open</strong> in Text Editor (Notepad or other you have specified) to verify. Enter sequence or copy and paste from another file.</td>
</tr>
</tbody>
</table>

*Table 39.1 Symbols indicating the resolution of the conversion process*
Other GCG Converter Operations

- Sort the columns by File Name, File Length or Type. Click on the column heading; repeat the click to reverse the sort.
- Click the Remove Selected or Remove All to remove all files from the list.
- Click the Info button, opening an Information box displaying the nucleotide content of the DNA molecule.
- Click the Edit button to open the text editor, where you can review or edit the file.

Importing into Vector NTI Database Explorer

In the Results section of the dialog box, the number of each type of molecule is displayed beneath the type symbols, DNA and Protein (Figure 39.23). To import converted molecules into Vector NTI, select the molecule file(s) in the Files to Convert section and double-click on the corresponding file type button in the Results section.
If you have selected more than one molecule to be imported, a Multiple Entry dialog box opens listed the molecules (Figure 39.24).

![Multiple Entry dialog box](image)

Figure 39.24 Multiple Entry dialog box

Click **Import** to import them into the Vector NTI database. In the dialog box that opens, specify the subset where they will be placed. You might want to create a new subset labeled **Converted from GCG**.

If you click the **Open** button, molecule display windows will be opened and tiled for each of the molecules on the list. You may see a message saying that not all of the restriction sites, and motifs can be opened for some of them due to space limitations.

Matrix Editor

All algorithms designed to evaluate pairwise sequence alignment are based on systems which rank aligned residues. Amino acids that are identical or similar in alignment score higher than those less similar. Matrices generated with these assigned scores are used to detect similarities between differing sequences. The most common of many different scoring systems are based on substitutions of amino acids in related proteins.

PAM and BLOSUM substitution matrices are objects stored in the Vector NTI Matrices folder. Customized matrices may also be devised and stored in Vector NTI for shared or private use. Global alignments of closely related sequences are used to derive the PAM matrices. Scores are based on extrapolated evolutionary differences between amino acids. The greater the numbers, the greater the evolutionary divergence. (Dayhoff et al. 1978) BLOSUM matrices are calculated from local, ungapped alignments of distantly related sequences. The matrix scores represent the minimum percent identity of the blocks used to construct the matrix; greater numbers represent lesser distances.(Henikoff and Henikoff 1993).

Matrix Editor allows inspection of and editing of matrices.

Matrices are used in two Vector NTI Advance applications, AlignX and AlignX Blocks. Matrices can be viewed in either program in the Alignment Setup dialog box, but editing can only take place in Matrix Editor.
To open the Matrix Editor, select **Start > Programs > Invitrogen > Vector NTI Suite 10 > Utilities > Matrix Editor**. The Matrix Editor that opens is initially empty (Figure 39.25).

Figure 39.25 Matrix Editor (initially empty)

Click the **Load Matrix** button to open the Matrices folder (Figure 39.26). Select the matrix you wish to edit/review and click **Open**.

Figure 39.26 Matrices folder

Figure 39.27 Matrix Editor (loaded)
The Matrix Editor (Figure 39.27) consists of two tabs: Score Matrix and Stats-Table.

The **Score Matrix** tab displays the name and order of the matrix residues. Use the scroll bars to view the entire matrix. The arrow buttons in the lower left of the Editor can be used to move the cursor vertically through the left-most column. To edit a cell, highlight a value and enter the new value (Figure 39.28).

![Figure 39.28 Editing a matrix cell](image)

The **Save Matrix** button is enabled only after you modify the name in the Name text box. This is to prevent any of the default matrices in Vector NTI from being altered.

Click the **Clear Matrix** button to remove the currently displayed matrix from the Editor.

The **Stats-Table** tab displays the Lambda and Kappa parameters used to calculate the statistical significance of high scoring segments or subalignments. On this tab, you can edit values in the same way you do on the Score Matrix tab. To save the edited matrix, you must return to the Score Matrix tab and modify the name, thus enabling the **Save Matrix** button. In the dialog box that opens, specify the file for storing the matrix and click the **Save** button.

### Oligo Importer

You can import oligo lists into the Vector NTI database using the Oligo Importer utility. The oligo list to be imported must be in one of three file formats for use with the Oligo Importer utility:

- **Comma delimited** format (.csv format)
- **Tab delimited** format
- **Semicolon delimited** format

The Oligo Importer utility will recognize three fields from the oligo list file:

- **Oligo Name**
- **Oligo Sequence**
- **Oligo Description**

The following is an example of an oligo list saved in .csv format from MS Excel and opened in Notepad for viewing (Figure 39.29):

![Figure 39.29 Oligo list saved in .csv format from MS Excel and opened in Notepad](image)

There are two steps in the oligo list import process:

- **Step 1**: Open the file in the Oligo Import (.scv file) utility and convert it to a VNTI oligo archive file.
- **Step 2**: Import the oligo archive file into the VNTI Database.

To import an oligo list file in comma, tab or semicolon delimited format, launch the Oligo Import utility by choosing **Start > Programs > Invitrogen > Vector NTI Suite 10 > Utilities > Oligo Import**.

The Oligo Import dialog box opens. Click the **Open** button in the upper right corner of the Oligo Import dialog box. In the Open File dialog box, locate the file you want to import, select it and click the **Open** button. The file, if in one of the accepted formats, opens in the Oligo Import dialog box (Figure 39.30):

![Figure 39.30 Oligo Import dialog box](image)

Select the delimiter type, if not already pre-selected. Press the column-heading bar over each data column and define the Name, Sequence and Description columns by choosing from the drop down list. All three columns **MUST** be defined even if they contain no data. For example, in the sample oligo file above, no description data exists, so the description column in the Oligo Importer contains no data, yet its column heading still must be defined or the file conversion cannot proceed.

Highlight the oligo rows for those oligos you wish to import (do not select rows that contain heading labels, if any). Click the **Import** button.

![Figure 39.31 The archive icon indicates successful file conversion](image)

If successful, the file is converted to a VNTI oligo archive file type and an icon for the archive file is displayed on the lower portion of the Oligo Import dialog box (Figure 39.31).

Double clicking on the archive icon opens the VNTI Database Explorer window. You can import the oligo archive into VNTI by dragging and dropping the icon from the Oligo Import dialog box onto the desired oligo subset in the Exploring Local VNTI Database window.
You can save the oligo archive file to your hard drive by dragging and dropping to the Desktop or to another location in Windows Explorer.

License Manager

Once you have installed Vector NTI Advance 10, you will need to license the application to be able to use it.

To satisfy the needs of users in different industrial, scientific or educational environments, Invitrogen has designed three types of Vector NTI licenses, as well as a Demo Mode option. These are all administered through the License Manager.

- **Static License**: Purchased by one user for installation on one computer
- **Dynamic License (DLS)**: A license that is installed on a server and issued by that server to client Vector Advance computers. DLS licenses are shared by a specified number of users or “seats,” with the number of users at any one time being limited to the number of “licenses” specified in the contract.
- **Trial License**: Allotted to a potential purchaser of Vector NTI software for a specified number of days, during which the user can review and use the software within certain limits.
- **Demo Mode**: When unlicensed, Vector applications operate in Demo Mode for the purpose of demonstrating the basic features of the software. Some functionality is disabled in Demo Mode.

When you open the Vector NTI Advance software, a checkmark icon, such as this ( ), at the bottom right corner on the Status Bar shows the current license status. Pause the cursor arrow over the button, and a pop-up label displays the license status.

- Green checkmark = active Static License
- Green, **blinking** checkmark = active Trial or Dynamic License
- Red, **blinking** X = the application is not licensed; running in Demo mode.

License Manager does not open automatically when you install Vector Advance on your computer. You must open License Manager manually. To open License Manager, select it from the **Start** menu: **Start > Programs > Invitrogen > Vector NTI Advance > License Manager** or click **Help > License** in the Database Explorer or a Vector NTI viewer window once you have opened it.

Launching License Manager

License Manager does not open automatically when you install Vector Advance on your computer. You must open License Manager manually. To open License Manager, select it from the **Start** menu: **Start > Programs > Invitrogen > Vector NTI 9 > License Manager** or click **Help > License** or **License Manager** in any Vector NTI Advance viewer once you have opened it.

License Manager Dialog Box

The License Manager consists of three tabs, the **Contact Us** tab, the **Personal** tab, and the **Applications** tab.
Contact Us Tab

The License Manager opens by default to the Contact Us tab (Figure 39.32). This tab summarizes your Vector software licensing agreement. Additionally, it provides information for upgrading your Vector application license and contacting Invitrogen.

![Figure 39.32 License Manager (Contact Us tab)](image)

Personal Tab

The Personal tab (Figure 39.33) provides text boxes for entering personal information. Once entered on this tab, when you click your license choice on the Applications tab, your entries are automatically entered on the license application.

![Figure 39.33 License Manager (Personal tab)](image)

Applications Tab

The Applications tab (Figure 39.34) indicates the type of License currently in effect for each Vector NTI Advance application, as well as for Vector Xpression and PathBlazer. If you have Vector software with a current license, the contract number displays in a cell to the right of the
corresponding application. For applications for which there is no current license, a drop-down menu displays in the cell to the right of the corresponding application.

```
Figure 39.34 License Manager (Applications tab)
```

- For a new installation or update of a previously unlicensed installation, License Manager opens in Demo mode for all applications.
- For Static, Dynamic and Trial licenses, if you are not licensing the entire software package using the same type of license, click in the license-type text box of the application for which you wish to specify a license. Click the down-arrow to extend the drop-down menu and select the appropriate license type. (Demo mode is the default selection.)

**Note:** If your active license type varies for different applications, this dialog box may have a combination of rows with drop-down menus and rows displaying contract numbers. Click the button appropriate for the license type you want to register. Each option is described in the following sections.

**Static License Dialog Box**

Static licenses can be obtained in either a personal or a corporate mode. To configure a static license, click the Static button at the bottom of the Applications tab (Figure 39.34). This opens the Static License dialog box (Figure 39.35).

```
Figure 39.35 Static License dialog box
```
For either type of static license, enter your name, organization, phone number and email address in the appropriate fields. This sets the user information in Vector NTI Advance.

**Note:** If you already entered your personal information on the Personal tab, it should appear here when you open this dialog box.

In the License # field, enter your Vector NTI Advance 10 static license number provided in the e-mail you received if you downloaded the software from the Invitrogen website, or in the letter that accompanied your CD ROM and/or manual.

- **Personal Static License:** You must have an internet connection for this step. Click the Apply button. The registration Key is returned by the TCP/IP protocol right after you press the Apply button, and it will be applied automatically.

  If the registration fails because of a missing connection to the Invitrogen licensing server, an appropriate message immediately displays. In such a case, you can contact Invitrogen Customer Support or Sales, providing them your computer’s hardware ID and your license number.

  Once you receive the registration key, enter the key in the Key text box of the Static License dialog box. Make sure the License Number is entered appropriately, and click Apply. If the Key matches your license number and computer hardware ID, the license is registered.

- **Corporate Static License:** No internet connection is needed nor is a Registration Key required. The License Manager on your client computer validates the Corporate License number; if it is valid, then your software is registered.

  Once you receive the registration key, enter the key in the Key text box of the Static License dialog box. Make sure the License Number is entered appropriately, and click Apply. If the Key matches your license number and computer hardware ID, the license is registered. No connection to the Internet is required in this case.

**Notes:**

- Once you have applied your static license, notice that the Applications tab reflects your static license status.

- If you want to reset your personal static license, (you cannot re-set a corporate static license) type Unregister in the License Number field and click Apply. You will be warned that you are trying to reset your static license and asked if you want to continue. If you answer Yes, the application will reset your license and will send proof of this operation to the Invitrogen server. If the connection to the server fails, you will receive notice of this.

**Dynamic License Dialog Box**

Technically, a Dynamic License is not a “license” mode, but registers permission to work in a shared DLS environment in license mode. To configure your Dynamic license, click the
Dynamic button at the bottom of the Applications tab (Figure 39.34). This opens the Dynamic License dialog box (Figure 39.36):

1. Enter your name, organization, phone number and email address in the appropriate fields. This sets the user information in Vector NTI Advance 10.

   **Note:** If you already entered your personal information on the Personal tab, it should appear here when you open this dialog box.

2. In the URL of DLS text box, enter the DLS server URL supplied by the DLS administrator at your site. If your DLS server requires a password, make sure the authentication settings are filled in appropriately.

3. Press the Internet Connection Settings button to configure your connection settings and to enter server proxy information, if a firewall is used at your site. Refer to the following section, Internet Connection Settings.

4. Press the Test Connection button to review the status of your connection. Refer to Testing the License Server Connection on this page.

5. Once you have configured the Dynamic License dialog box parameters, to set all Vector applications to Dynamic License, press the Set For All Applications button. Once you do this, when you close this dialog box, the Applications tab now shows Dynamic License for all applications.

   **Note:** When you set Dynamic licenses for all applications, this operation only applies for those applications for which you do not have a Static License.

   Press the Apply button to execute the dynamic license configuration.
**Internet Connection Settings**

For a Dynamic License, press the **Internet Connection Settings** button in the Dynamic License Server Connection dialog box. This opens the Internet Settings dialog box where you can alter your proxy settings (Figure 39.37):

![Image of Internet Settings dialog box](image)

**Figure 39.37 Internet Settings dialog box**

The Internet Settings dialog box allows you to set your connection parameters. If the **Use Internet Explorer settings** button is selected, License Manager will attempt to make the connection using your default settings. If default detection is not successful, you can either choose the **Direct connection** button if you do not have a proxy or choose the **Use proxy server** button and specify the proxy name, port and password information.

Press the **OK** button to return to the Dynamic License Server Connection Tester dialog box. See the Internet Connection Settings section on page 655 for more information.

**Testing the License Server Connection**

In the Dynamic License dialog box, press the **Test Connection** button to review the status of your connection. This opens the Server Connection Tester dialog box (Figure 39.38).

![Image of Dynamic License Server Connection Tester dialog box](image)

**Figure 39.38 Dynamic License Server Connection Tester dialog box**

The status of the connection displays in the right-hand panel.

- If the server requires a password, it must be entered into the corresponding text box in this dialog box.
- If you want to alter your proxy settings, press the **Internet Connection Settings** button (see next section).
- Once the settings are reconfigured, press the **Connect** button to test the connection using the new settings.

Once you have configured the **Dynamic License** dialog box parameters, to set all Vector applications to Dynamic License, press the **Set For All Applications** button. Once you do this, when you close this dialog box, the Applications tab now shows Dynamic License for all applications.
**Note:** When you set Dynamic licenses for all applications, this operation only applies for those applications for which you do not have a Static License.

Press the **Apply** button to execute the dynamic license configuration.

**Trial License Dialog Box**

To configure a trial license, click the **Trial** button at the bottom of the **Applications** tab of License Manager (Figure 39.34). This opens the Trial License dialog box (Figure 39.39):

![Trial License dialog box in License Manager](image)

As described in the Trial License dialog box, you must generate an e-mail addressed to bioinfo-trial@invitrogen.com and include your name, the name of your organization, phone number, and email address.

**Notes:**
- The e-mail will not be generated automatically from this dialog box.
- You will receive a trial license number in response from an Invitrogen contact person. Return to this dialog box and enter the trial license number in the text box. Click the **Apply** button to complete the trial license registration.

When you set Trial Licenses for all applications, this operation only applies for those applications for which you do not have a Static License.

**Vector NTI System Information**

The Vector NTI System Information application collects data about Vector NTI and the system on which Vector NTI is installed and consolidates this information into one file. The System Information file is helpful in the diagnosis and resolution of system-related Vector NTI problems. In such cases, the Invitrogen Customer Support team may request that you run the Vector NTI System Information utility and e-mail them the resulting file.

The Vector NTI System Information utility program’s file name is `vntisysinfo.exe` and it is located in the Vector NTI Suite 10 folder on your hard drive. To run the utility, navigate to the Vector NTI Suite 10 folder and double-click the `vntisysinfo.exe` application in the Vector NTI program files on your hard drive. The Vector NTI System Information window displays (Figure 39.40):

![Vector NTI System Information window](image)
Click the **Collect Info** button to initiate the system information gathering process. The Settings dialog box displays (Figure 39.41), allowing you to select the categories of information the Vector NTI System Information utility gathers:

![Figure 39.41 Vector NTI System Information Settings dialog box](image)

You can choose from three preset data gathering selection options:

- **Required** – the minimum required system information is gathered.
- **Recommended** – the minimum required information is gathered along with additional system information that is often useful.
- **All** – gathers system information in all listed categories.

You can also select or deselect individual categories manually by clicking the checkbox next to the category.

Once you have configured the Settings appropriately, click the **Proceed** button to begin system data collection. When the utility has completed collecting the system data, a data report appears in the Vector NTI System Information dialog box (Figure 39.42):

![Figure 39.42 Vector NTI System Information showing collected system data report](image)

Expand the folders to show all the collected data and click the **Copy All** button to copy the report to the clipboard for pasting into another document or press the **Send** button to open your
e-mail program so you can send the Vector NTI System Information report to Invitrogen Customer Support (bioinfotrial@invitrogen.com).
As you use Vector NTI Advance 10, if your questions are not answered in the Vector NTI Advance 10 User's Manual, you may find it helpful to contact Invitrogen Customer Support. You can also refer to the Invitrogen Web site for frequently updated information regarding your Vector NTI Advance 10.

Invitrogen Technical Support

Invitrogen's free technical support for Vector NTI software is available exclusively through the web. For more information, check out the eSupport section of the Vector NTI user community at http://www.invitrogen.com/vectorNTIcommunity.

To obtain personalized technical support by telephone or email, you must have an annual support contract. Academic and government users may purchase an Academic Vector NTI Support Contract through the User Community. Commercial users may purchase an Advanced Support Contract by contacting Invitrogen at bioinfosales@invitrogen.com (for the Americas) or eurobioinfoorders@invitrogen.com (for European/Middle Eastern/African/Asian Pacific customers).

For paid support, use the following contacts:

United States

Phone: 240-379-4240 or 877-357-3114 (Toll-free, U.S.)
E-mail: bioinfosupport@invitrogen.com

Europe, Middle East, Africa, Asian Pacific

Phone: +44 (0) 141 814 6350
Email: eurobioinfosupport@invitrogen.com

The Invitrogen Technical Support team may request that you run the Vector NTI System Information utility and email them the resulting file. Vector NTI System Information is a utility application, which gathers data about the system on which Vector NTI is installed and consolidates that...
information into one file. The system information file is helpful in the diagnosis and resolution of system-related Vector NTI problems.

**Invitrogen Web Site**

Click the Invitrogen toolbar button (in Vector NTI Advance to access the Invitrogen web site, www.invitrogen.com.
3D-Mol

Connolly Surface Calculation


Varshney Surface Calculation

- Amitabh Varshney, Frederick P. Brooks, Jr, William V. Wright. Interactive Visualization of Weighted Three-dimensional Alpha Hulls. Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3175.
- Dissertation of Amitabh Varshney. Hierarchical Geometric Approximations. (Under the direction of Professor Frederick P. Brooks, Jr.)

AlignX

Basic Alignment Algorithms

Multiple Alignment

The alignment is created using the Clustal W algorithm (Nucleic Acid Research, 22 (22): 4673-4680, 1994). Briefly, a crude similarity between all pairs of sequences is calculated, called a “Parities alignment”. These scores are then used to calculate a "guide tree" or dendrogram,
which tells the multiple alignment stage the order in which to align the sequences for the final multiple alignment. Having calculated the dendrogram, the sequences are aligned in larger and larger groups until the entire sequences are incorporated in the final alignment.

In the AlignX modified Clustal W algorithm, the user designates the scoring matrix used by the algorithm rather than the algorithm making the choice as is done in a traditional Clustal W alignment. (Saul B. Needleman and Christian D. Wunsch. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of two proteins. Mol. Biol. (1970) 48, 443-453.)

**Profile Alignment**

The term “profile alignment” describes the alignment of 2 alignments. The method is a simple extension of the profile method of Gribskov, et. al. (Gribskov, M.; McLachlan, A.D. and Eisenberg, D. (1987) Profile analysis: detection of distantly related proteins. PNAS USA 84, 4355-4358) for aligning a sequence with an alignment.

**Substitution Matrices**

**PAM**


**BLOSUM**


**Phylogenetic Tree**


**Alignments and Similarity**


**BioAnnotator**

**DNA/RNA Analyses Citations**

- Susan M. Freier et al. PNAS 83, 9373-9377 Improved free-energy parameters for predictions of RNA duplex stability.
- Kenneth J. Breslauer et al. PNAS 83, 3746-3750 Predicting DNA duplex stability from the base sequence.
Protein Analyses (ProtScale Analyses)


Properties Plots

- Cowan, R., and Whittaker, R.G. Hydrophilici ty indices @ pH 3.4 determined by HPLC. Peptide Research 3, 75-80.


**ContigExpress**

**Contig Assembly**


**BLAST**

- Tatusov and Lipman, in preparation.

**Miscellaneous**

**Molar Extinction Coefficient**

The molar extinction coefficient for a Protein Molecule reported in the Analysis folder in the Protein Molecule display window is calculated according to the following reference: Gill and von Hippel, Anal. Biochem. 182, 319-326 (1989). The calculation assumes the protein is denatured under the following conditions: 6.0 M guanidium hydrochloride, 0.02 M phosphate buffer, pH 6.5.

**Oligo Analysis**


PCR Analysis


General Bioinformatics


Biostatistics


REBASE

APPENDIX C

ASCII FORMAT; IUB CODES (IUPAC CODES)

Format for ASCII Sequence Files

An ASCII sequence file must obey the following rules:

- It must be a plain (ASCII) text file (not a file in WordPerfect, or MS Word formats).
- The file must contain the nucleotide (amino acid) sequence arranged in lines. Each line may contain the following:
  - Nucleotide (amino acid) symbols and white space, or
  - A number followed by white space and nucleotide (amino acid) symbols (therefore, similar to GenBank format), in which case the number will be ignored, or
  - A number only, in which case the number will be interpreted as a block of unknown nucleotides (amino acids) of the corresponding length.

IUB Formats recognized by Vector NTI

The following characters, defined by the International Union of Biochemistry (IUB), are used to represent nucleotides throughout Vector NTI:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>adenine</td>
</tr>
<tr>
<td>T</td>
<td>thymine</td>
</tr>
<tr>
<td>C</td>
<td>cytosine</td>
</tr>
<tr>
<td>G</td>
<td>guanine</td>
</tr>
<tr>
<td>R</td>
<td>purine(A or G)</td>
</tr>
<tr>
<td>Y</td>
<td>pyrimidine(C or T)</td>
</tr>
<tr>
<td>W</td>
<td>A or T</td>
</tr>
</tbody>
</table>

Table C. 1 IUB Formats for nucleotides
Amino acids are represented by standard 1- (or 3-) letter codes:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>C or G</td>
</tr>
<tr>
<td>M</td>
<td>A or C</td>
</tr>
<tr>
<td>K</td>
<td>T or G</td>
</tr>
<tr>
<td>B</td>
<td>C, G, or T</td>
</tr>
<tr>
<td>D</td>
<td>T, G, or A</td>
</tr>
<tr>
<td>H</td>
<td>C, A, or T</td>
</tr>
<tr>
<td>V</td>
<td>C, G, or A</td>
</tr>
<tr>
<td>N</td>
<td>unknown nucleotide</td>
</tr>
</tbody>
</table>

Table C. 1 IUB Formats for nucleotides

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>Glu</td>
</tr>
<tr>
<td>G</td>
<td>Gly</td>
</tr>
<tr>
<td>H</td>
<td>His</td>
</tr>
<tr>
<td>I</td>
<td>Ile</td>
</tr>
<tr>
<td>L</td>
<td>Leu</td>
</tr>
<tr>
<td>K</td>
<td>Lys</td>
</tr>
<tr>
<td>M</td>
<td>Met</td>
</tr>
<tr>
<td>F</td>
<td>Phe</td>
</tr>
<tr>
<td>P</td>
<td>Pro</td>
</tr>
<tr>
<td>S</td>
<td>Ser</td>
</tr>
<tr>
<td>T</td>
<td>Thr</td>
</tr>
<tr>
<td>W</td>
<td>Trp</td>
</tr>
<tr>
<td>Y</td>
<td>Tyr</td>
</tr>
<tr>
<td>V</td>
<td>Val</td>
</tr>
<tr>
<td>B</td>
<td>Asx</td>
</tr>
<tr>
<td>Z</td>
<td>Glx</td>
</tr>
<tr>
<td>X</td>
<td>???</td>
</tr>
</tbody>
</table>

Table C. 2 IUB Formats for amino acids
Overview of Parent-Descendant Relationships

Vector NTI tracks and manages parent-descendant connections between constructed molecules.

Consider the case of the two constructed DNA molecules, NEWMOL1 and NEWMOL2:

- NEWMOL1 is constructed from two fragments, a fragment of molecule A and a fragment of molecule B.
- NEWMOL2 is constructed from two fragments, a fragment of NEWMOL1 and a fragment of molecule C.

In this case, we have the following parent-descendant connections between A, B, C, NEWMOL1 and NEWMOL2:

- A and B are direct ancestors or parents of NEWMOL1
- NEWMOL1 and C are direct ancestors or parents of NEWMOL2
- A, B, NEWMOL1, and C are ancestors of NEWMOL2
- NEWMOL1 is the direct descendant of A and B
- NEWMOL2 is the direct descendant of NEWMOL1 and C
- NEWMOL2 is the descendant of A and B

Figure D.1 Parent-descendant relationships
Vector NTI also manages the following molecule relationships:

Protein molecules constructed by translating a coding region of a DNA molecule are “direct descendants” of the DNA molecule from which they are translated.

Molecules constructed from PCR analysis of a region of a DNA molecule are direct descendants of that DNA molecule.

Parent-descendant connection data is useful in tracking constructs in your database and for automatically passing changes in parent molecules to descendant molecules.

**Viewing Parent-Descendant Relationships**

Parent-descendant relationships for a molecule can be viewed in tree form in the molecule Properties dialog box. You can launch the Properties dialog box for a molecule from the Exploring Vector NTI Database window by highlighting the molecule listing of interest in the List Pane of the database window and either choosing <molecule type> > Properties from the menu or by right clicking on the molecule and choosing Properties from the shortcut menu.

Choose the Parents tab to view the molecule’s parents (Figure D.2):

![Figure D.2 Parents tab on Molecule Properties dialog box](image)

Choose the Descendants tab to view the molecule’s descendants (Figure D.3):
Managing Parent-Descendant Relationships

Vector NTI tracks and maintains the integrity of parent-descendant molecule relationships. This has obvious implications in the following situations:

- If you change the sequence of a descendant molecule
- If you change the sequence of a parent molecule
- If you delete a parent molecule from the database

To ensure the integrity of parent-descendant relationships, Vector NTI does not allow modification or deletion of such molecules without your approval and it makes the appropriate adjustments to the affected parent-descendant connections when any of the above actions are performed.

Modifying a Descendant

If you attempt to modify a child molecule, the following warning message is issued (Figure D.4):

Important: Once parent-descendant relationships have been disconnected, they CANNOT be reconnected.
Modifying a Parent

If you attempt to modify a parent molecule, when you try to save the altered molecule to the database under the original name, the following warning appears (Figure D.5):

![Parent molecule modification warning](image)

**Basic options:**

- Recalculate Descendants: Changes the parent and all its descendants to incorporate the edit.
- Disconnect Descendants: Changes the parent only. Descendants remain unchanged, but are disconnected from their parents and become Basic molecules in the database.

**If a Descendant Cannot be Recalculated:**

There may be instances where it is impossible for VNTI to recalculate a descendant molecule when you change one of its parents. This typically happens when one of the parent molecules is absent from the database.

For example, descendant molecules can be imported into the local VNTI database from a shared database or VNTI molecule archive with the links to their ancestor molecules preserved, but without the ancestor molecules being present. In the case of a descendant molecule that is imported along with the first parent, but with only a link to the second parent, if you modify the first parent, VNTI is unable to recalculate the descendant to include the edits, because the second parent is not present in the database.

In such cases, you have the following choices:

- Leave Unprocessed: Descendant becomes an “unprocessed” molecule in the database; unprocessed molecules cannot be opened in a Molecule Viewing window, but they can be reconstructed (once all missing parent molecules have been imported into the database).
- Disconnect: Descendants remain unchanged, but are disconnected from their parents and become Basic molecules in the database.

**Important:** Once parent-descendant relationships have been disconnected, they CANNOT be reconnected.
Deleting a Parent from the Database

When deleting a parent molecule from the database, you have two choices:

- **Delete from Database**: Deletes only the parent molecule. All descendant molecules become Basic molecules in the database. The following warning is issued when the Delete from Database command is executed (Figure D.6):

![Figure D.6 Warning issued when deleting a parent from the database](image)

- **Delete with Descendants from DB**: Deletes the parent molecule and all its descendants from the database. When this command is executed, the following warning is issued (Figure D.7):

![Figure D.7 Warning issued when deleting with descendants](image)

**Important**: Once the OK button on the warning dialog box is pressed, depending on the delete command chosen, the parent molecule or the parent and all its descendants will be immediately deleted from the Database and all subsets. Deleted molecules CANNOT be “undeleted”.

Deleting a Descendant from the Database

When a descendant that is not a parent to any other molecule (for example, NEWMOL2 in the example above, see Figure D.1) is deleted, the only connection broken is the connection between the deleted molecule and its parent. No other connections are affected. Therefore, the standard molecule deletion warning is issued (Figure D.8):

![Figure D.8 Standard molecule deletion warning](image)

The warning gives no indication that parent-descendant connections will be affected, because the only connection that is broken is that between the deleted molecule and its parent.

**Important**: Once the OK button is pressed, the molecule is deleted from the database and all subsets. Deleted molecules CANNOT be “undeleted”.

Appendix E

TM Calculations in Vector NTI

General Information

Vector NTI calculates and reports two different melting temperatures for DNA/RNA oligonucleotides, Thermodynamic Tm (Therm. Tm) and %GC Tm (Figure E.1):

Usefulness of Thermodynamic Tm Versus %GC Tm

Vector NTI reports both the Thermodynamic and %GC Tm values, regardless of the length of the oligo (oligos are limited to 1000 characters maximum length in Vector NTI 9). However, generally only one of the reported TM values should be considered useful, depending on the length of the oligo as follows:
- Therm. Tm – useful for oligos that are greater than about 7-10 residues and less than about 35 residues long
- %GC Tm – useful for oligos greater than about 35 residues long

Note: For oligos that are 7-10 residues or shorter (cutoff length depends on the base content of the particular oligo being analyzed), Vector NTI reports a Therm. Tm value of zero.

Effects of Primer (Probe) and Salt Concentration on Tm Calculations

Tm calculations are highly dependent on primer and salt concentrations; varying these concentrations can greatly affect the Tm for any given primer. Therefore, it is important that you adjust the primer and salt concentrations appropriately so that accurate Tm values are generated.

Note: In Vector NTI, the default parameters for primer and salt concentration are 250 pM and 50 mM, respectively, for calculating Tm values. Other Tm calculators commonly use a default probe concentration of 50 nM. Because of this, Vector NTI default parameter Tm values may not correspond to the default Tm values calculated using other programs. Before comparing Vector NTI Tm values with those generated by other Tm calculators, make sure that the parameters are adjusted appropriately.

%GC Tm Calculation

The %GC Tm calculation does not rely on the thermodynamic properties of the oligo (i.e. dHo, dSo and dG° values). The formula for %GC Tm is as follows:

\[
Tm = 81.5 + 16.6(\log[Na^+]) + 0.41(\%GC) – 675/probe length
\]

Note: [Na+] is in molar units.

Example: For oligo GTGCGAGGCAGCTGCGGTAA at 50mM salt:

\[
Tm = 81.5 + 16.6(\log(0.05)) + 0.41(65) – 675/20
\]

\[
= 81.5 + 16.6(–1.30) + 26.65 – 33.75
\]

\[
= 81.5 – 21.58 + 26.65 – 33.75
\]

\[
= 52.82 ^\circ C
\]

Thermodynamic Tm Calculation

The Thermodynamic Tm calculation is based on the Nearest Neighbor theory of DNA/RNA duplex stability. Briefly, this theory states that the overall duplex stability (and, hence, the melting temperature) of an oligonucleotide can be predicted from the primary sequence based on the relative stability and temperature-dependent behavior of every dinucleotide pair in the oligo. In practice, enthalpy (dH°) and free energy (dG°) values for each of the 10 possible Watson-Crick DNA pairwise interactions are used to calculate pairwise entropy (dS°) values via the following standard equation:

\[
dG° = dH° - TdS°
\]

Note: T is temperature in °K.

The pairwise dH° and dS° values are then summed to calculate overall values for the oligo under consideration. The overall values are used in the following formula to calculate the Thermodynamic Tm:

\[
\text{Therm. Tm} = dH° - 273.15 + 16.6(\log[Na^+]) dS° + dSo° + R(ln(c/4))
\]
Notes:
- dSo° is the entropy associated with helix initiation (-10.8 cal/mol per °K).
- R is the Universal Gas Constant (1.987 cal/mol per °K).
- c is the concentration of the probe, in molar units.
- The factor −273.15 corrects for absolute temperature so that the final Tm is in °C.

The pairwise dH° and dS° values for DNA used in VNTI are taken from reference 2. Those values, along with the corresponding dG° values at 25°C, appear in the following table (Table E.1):

<table>
<thead>
<tr>
<th>Interaction</th>
<th>dH° kcal/mol</th>
<th>dS° cal/mol per °K</th>
<th>dG° kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA/TT</td>
<td>-9.1</td>
<td>-24.0</td>
<td>-1.9</td>
</tr>
<tr>
<td>AT/TA</td>
<td>-8.6</td>
<td>-23.9</td>
<td>-1.5</td>
</tr>
<tr>
<td>TA/AT</td>
<td>-6.0</td>
<td>-16.9</td>
<td>-1.0</td>
</tr>
<tr>
<td>CA/GT</td>
<td>-5.8</td>
<td>-12.9</td>
<td>-2.0</td>
</tr>
<tr>
<td>GT/CA</td>
<td>-6.5</td>
<td>-17.3</td>
<td>-1.3</td>
</tr>
<tr>
<td>CT/GA</td>
<td>-7.8</td>
<td>-20.8</td>
<td>-1.6</td>
</tr>
<tr>
<td>GA/CT</td>
<td>-5.6</td>
<td>-13.5</td>
<td>-1.6</td>
</tr>
<tr>
<td>CG/GC</td>
<td>-11.9</td>
<td>-27.8</td>
<td>-3.6</td>
</tr>
<tr>
<td>GC/CG</td>
<td>-11.1</td>
<td>-26.7</td>
<td>-3.1</td>
</tr>
<tr>
<td>GG/CC</td>
<td>-11.0</td>
<td>-26.6</td>
<td>-3.1</td>
</tr>
<tr>
<td>XX/XX</td>
<td>-6.0</td>
<td>-16.9</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

Table E.1  DNA Nearest Neighbor thermodynamics

Notes:
- All values refer to the disruption of a duplex at 1 M NaCl, 25°C and pH 7.
- The units for dH° and dG° are kcal/mol of interaction, whereas those for dS° are cal/°K per mol of interaction.

Example: The oligo 5'-GTGCGAGGCAGCTGCGGTAA-3' is parsed as follows (Figure E.2):

![Figure E.2 Parsed Oligo](image)

Total dH° = -164.7 kcal/mol (Figure E.2)

The total dS° reported by VNTI (Figure E.2) is the sum of the pairwise values above and the entropy associated with helix initiation (dSo°). Thus, for the example oligo above:

Total dS° = -405.7 + (-10.8) = -416.5 cal/mol per °K

The total dG° (Figure E.2) is the sum of the pairwise dG° values for the oligo plus a helix initiation free energy term (dGo°) that is added to better reflect experimentally determined free energy values for tested oligos. The value of the helix initiation free energy term (dGo°) depends on the base composition of the oligo2 as follows:
- +5.0 kcal/mol for oligos containing any G-C base pairs
- +6.0 kcal/mol for oligos composed exclusively of A-T base pairs
Therefore, for the example oligo:

Total dGo = -43.7 kcal/mol (sum of the pairwise dGo values) + 5 kcal/mol (free energy term)

= -38.7 kcal/mol

The 3’ End dG° is calculated using the number of 3’ pairwise dG° values specified in the 3’ End Length (bp) box, and is not further adjusted (Figure E.3).

Using Vector NTI’s default probe and salt concentrations (250 pM and 50 mM, respectively) and the values for dH° and dS° calculated above, Therm. Tm can be calculated as follows:

\[
\text{ThermTm} = \frac{dH^°}{dS^° + dS_{o}^° + R \left( \ln \left( \frac{C}{4} \right) \right)} - 273.15 + 16.6 \left( \log Na \right)
\]

\[
... = \frac{-164.7}{-0.4165 + (0.001987) \left( \ln \left( \frac{250 \times 10^{-12}}{4} \right) \right)} - 273.15 + 16.6 \left( \log 0.05 \right)
\]

\[
... = \frac{-164.7}{-0.4165 + (0.001987)(-23.49)} - 273.15 + 16.6(-1.301)
\]

\[
... = \frac{-164.7}{(-0.4165) - 0.04667} + (-273.15) - 21.60
\]

\[
... = \frac{-164.7}{-4632} - 294.75
\]

\[
... = 355.57 - 294.75
\]

\[
... = 60.82°C
\]
Vector NTI adjusts the %GC and Therm. Tm values accordingly, based on the input formamide concentration (Figure E.4):

Figure E.4 Oligo Analysis dialog box showing effects of formamide on Tm

Oligos Containing IUB Ambiguity Characters

Vector NTI can analyze oligos that contain IUB nucleotide ambiguity characters (i.e. R, Y, W, S, M, K, B, D, H, V and N – See Appendix C). In the case of ambiguity characters, Vector NTI uses average pairwise $dH^\circ$ and $dS^\circ$ values for calculating the Tm.

For example, for the dinucleotide pair CB, Vector NTI averages the CC, CG and CT thermodynamic parameters (Table E.1) to obtain average pairwise $dH^\circ$ and $dS^\circ$ values for CB. It then sums the average pairwise thermodynamic parameters and calculates the Therm. Tm values according to the equation described above (see Thermodynamic Tm Calculation, page 676).

In the case of %GC Tm, Vector NTI applies the appropriate %GC contribution represented by each ambiguity symbol to the standard %GC Tm formula (see %GC Tm Calculation, page 676). For example, a B ambiguity symbol contributes only two-thirds the amount of a G or C residue to overall GC content.

RNA Oligos

RNA oligos use a different set of pairwise thermodynamic values than DNA oligos. Pairwise thermodynamic values for RNA are summarized in the following table (Table E.2):

<table>
<thead>
<tr>
<th>Interaction</th>
<th>$dH^\circ$ kcal/mol</th>
<th>$dS^\circ$ cal/mol/K</th>
<th>$dG^\circ$ kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA/UU</td>
<td>-6.6</td>
<td>-18.4</td>
<td>-1.1</td>
</tr>
<tr>
<td>AU/UA</td>
<td>-5.7</td>
<td>-15.5</td>
<td>-1.1</td>
</tr>
<tr>
<td>UA/AU</td>
<td>-8.1</td>
<td>-22.6</td>
<td>-1.4</td>
</tr>
<tr>
<td>CA/GU</td>
<td>-10.5</td>
<td>-27.8</td>
<td>-2.2</td>
</tr>
<tr>
<td>GU/CA</td>
<td>-10.2</td>
<td>-26.2</td>
<td>-2.4</td>
</tr>
<tr>
<td>CU/GA</td>
<td>-7.6</td>
<td>-19.2</td>
<td>-1.9</td>
</tr>
</tbody>
</table>

Table E.2 RNA Nearest Neighbor thermodynamics
Notes:

- All values refer to the disruption of a duplex at 1 M NaCl, 25°C, and pH 7.
- The units for $dH^o$ and $dG^o$ are kcal/mol of interaction, whereas those for $dS^o$ are cal/°K per mol of interaction.
- The $dS^o$ value for RNA oligos is adjusted by $-10.8$ cal/°K per mol to reflect the entropy associated with helix initiation, as it is for DNA oligos.
- The $dG^o$ value is adjusted by $+3.4$ kcal/mol to account for helix initiation. Note that this adjustment is NOT dependent on the base composition of the RNA oligo as it is for DNA oligos (see see Thermodynamic Tm Calculation, page 676).

Primer/Probe Tm, TaOpt and Similarity Calculations

For oligos designed using Vector NTI's PCR Primers, Sequencing Primers and Hybridization Probes features, the oligo Tm, product TaOpt and oligo percent binding similarity are reported in the Text Pane of the Molecule Viewing window.

Primer/Probe Tm Values

Tms for designed primers/probes are reported in Vector NTI:

- Therm. Tm is reported if the oligo is less than or equal to 35 residues
- %GC Tm is reported if the oligo is 36 residues or greater

For PCR products, Vector NTI reports the %GC; the assumption being that the majority of PCR products are larger than 35 residues.

TaOpt Values

The PCR product TaOpt (optimal annealing temperature for amplification of the fragment) in °C is calculated using the following formula:

$$TaOPT = ((0.3)Tm\text{primer} + (0.7)Tm\text{product} – 14.9) °C$$

Notes:

- $Tm\text{primer}$ is the Tm of the less stable primer of the pair
- $Tm\text{product}$ is the Tm of the PCR product

Primer/Probe Similarity Values

When designing PCR, sequencing or hybridization primers, VNTI reports the overall similarity in percent of an oligo to its binding site based on the oligo’s nucleotide composition. For oligos containing IUB ambiguity symbols, three similarity values are reported:

- Minimum Similarity
- Maximum Similarity
- Average Similarity

For Minimum Similarity, all ambiguities are classed as complete mismatches (i.e., they are assigned values of 0 at each position). For example, a 20mer containing 2 Rs and 2 Ns has a Minimum Similarity of 80%.

<table>
<thead>
<tr>
<th>GA/CU</th>
<th>-13.3</th>
<th>-35.5</th>
<th>-2.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG/GC</td>
<td>-8.0</td>
<td>-19.4</td>
<td>-2.2</td>
</tr>
<tr>
<td>GC/CG</td>
<td>-14.2</td>
<td>-34.9</td>
<td>-3.8</td>
</tr>
<tr>
<td>GG/CC</td>
<td>-12.2</td>
<td>-29.7</td>
<td>-3.3</td>
</tr>
<tr>
<td>XX/XX</td>
<td>-6.0</td>
<td>-16.9</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

Table E.2  RNA Nearest Neighbor thermodynamics (continued)
For Maximum similarity, all ambiguities are considered identical to their cognate nucleotides (i.e. they are assigned values of 1 at each position), so the Maximum Similarity is always 100%.

For Average similarity, Vector NTI weights each ambiguous nucleotide depending on whether it represents 2, 3, or 4 possible nucleotides. For example, Ns have a score of 0.25, Rs of 0.5, Bs of 0.33, etc. Therefore, for the 20mer described above, the Average Similarity is 85%.

References

Introduction

This appendix describes regular expression syntax, which can be used to perform sequence searches with the GenomBench Pattern Search Tool. A regular expression describes a pattern or sequence of adjacent characters and is written in a specific syntax. A brief description of regular expressions is given here based on the discussion in the book *Unix Power Tools*\(^1\). There are also many other printed and online sources where you can obtain a more in depth discussion about regular expressions.

There are two main types of regular expressions: simple and extended. Extended regular expressions provide some extra functionality over simple regular expressions by adding the ability to group search patterns, use the OR operator, and use several extra modifying characters. Most utilities such as *grep* understand simple regular expressions. The GenomBench Pattern Search Tool understands extended regular expressions. Other utilities that understand extended regular expressions are *egrep* and *awk*.

Parts of a Regular Expression

There are several main parts to a regular expression:

- Anchors: specify the position of a pattern relative to a line of text
- Character sets: match one or more characters in a single position
- Modifiers: specify the number of times the previous character set is repeated
- Grouping and pattern memory: groups and stores patterns in memory for later use

A regular expression can also contain special characters or 'metacharacters' that change the meaning of characters before or after the metacharacter. Metacharacters also allow for expanding and limiting possible matches. An example of a metacharacter is the dot ".", which matches any character. It is similar to a wildcard. For example, the simple expression `.A` matches any single character that is followed by the letter A.

---

Anchors

Anchors search for a pattern that is at one end of a line or the other. The caret '^^' is the starting anchor and the dollar sign '$' is the ending anchor. When a ^ is placed at the beginning of a regular expression, the beginning of a line is searched for the pattern. For example, the expression ^^AATT matches all lines that start with the string AATT. When a $ is placed at the end of a regular expression, the end of the line is searched for the pattern. For example, the expression AATT$ matches all lines that end in the string AATT.

Character Sets

Characters are matched as character sets where the simplest character set is a single character. For example, the regular expression **CG** consists of two character sets: **C** and **G**. Any line that contains the string **CG** is matched.

Square brackets [ ] are metacharacters that are used with character sets and allow a search for specific characters inside the brackets. Arbitrary characters are specified by using the syntax **[0-9]** to indicate any digit, **[a-z]** to indicate any lower case letter, and **[A-Z]** to indicate any upper case letter. A hyphen between characters in an ascending range specifies a range of characters. Character sets can also be concatenated. For example, **[0-9]** **[a-zA-Z]** matches any digit and then any letter.

Modifiers

Modifiers are used to specify the number of times a character set in a pattern should be searched. The asterisk '*' is a metacharacter that matches zero or more occurrences of a character set. For example, **[A]** * matches zero or more occurrences of the character A and **[A] [A] * matches two or more occurrences of the character A. Two other modifiers are added by extended regular expression syntax. A question mark '?' matches zero or one occurrences and a plus sign '+' matches one or more.

Grouping and Pattern Memory

Extended regular expression syntax also includes several additional characters that provide the ability to match a choice of patterns and then "remember" it. The important characters for grouping and pattern memory are parentheses '(' and ')'. Parentheses group expressions and are used to mark part of a pattern. The pipe symbol '|' acts as an OR operator within a grouping. For example, the expression

(AAAA|TTTT) TACG ( [CCCC | GGGG] )

matches the string TACG either preceded by the string AAAA or TTTT and followed by the string CCCC or GGGG. With extended regular expressions you can also use the modifiers '*', '?', and '+' within parentheses.

A backslash followed by a digit (for example, \1) is used to indicate that the value of an expression stored in pattern memory (that is, an expression in parentheses) should be substituted. In a replacement string, \0 is replaced by the entire matched string and \\1 through \\9 is replaced by the corresponding regular subexpression in a sequence of subexpressions. A backslash followed by a digit can also be used to repeat a pattern. For example, to search for two of the same characters, use **([A-Z]) \1**.

Subexpressions or nested expressions are also defined by parentheses. When there are nested parentheses, the order of subexpressions is defined by the order of open parentheses. If a subexpression matches several times, its value is defined by the last match. If a subexpression does not match, its value is an empty string. The following rules cause a subexpression to not match:

- It is suffixed by '**', '?' or '{0,m}' and the match count is zero.
- It is one of a choice of subexpressions separated by '|' and other subexpression variant matches.
- It is contained in another subexpression and the containing subexpression does not match.

An example of a nested expression is:

```
((one|many +([0-9]*)) *; *)*
```

The following table shows what the actual match is when different digits are used with the backslash and the expression is applied to the value:

```
many 20; one;
```

<table>
<thead>
<tr>
<th>When this digit is used...</th>
<th>This subexpression matches...</th>
<th>Actual match value is...</th>
</tr>
</thead>
<tbody>
<tr>
<td>\0</td>
<td>whole expression</td>
<td>many 20; one;</td>
</tr>
<tr>
<td>\1</td>
<td>(one</td>
<td>many +([0-9]*)) *; *</td>
</tr>
<tr>
<td>\2</td>
<td>one</td>
<td>many +([0-9]*)</td>
</tr>
<tr>
<td>\3</td>
<td>[0-9]*</td>
<td>20</td>
</tr>
</tbody>
</table>

Table F.1 Example of subexpression matches when different digits are used with '\'

**Metacharacter Reference**

The following table lists valid metacharacters that can be used to form extended regular expressions with the GenomBench Pattern Search Tool.

<table>
<thead>
<tr>
<th>Metacharacter</th>
<th>Function</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>^</td>
<td>Matches the following regular expression at the beginning of a line. To match a literal ^ at the start of a line, use the \ character.</td>
<td>^AATT: The string AATT at the start of a line AATT^: The string AATT^ anywhere on the line ^: A ^ at the start of a line</td>
</tr>
<tr>
<td>$</td>
<td>Matches the preceding regular expression at the end of a line. To match a literal $ at the end of a line, use the \ character.</td>
<td>$: An A at the end of a line $A: A $A anywhere on the line $$: A $ at the end of a line</td>
</tr>
<tr>
<td>.</td>
<td>Matches any single character except the new line character</td>
<td>.2: Any character following A such as ACT, AGA, AAA...CCC</td>
</tr>
<tr>
<td>\</td>
<td>Escapes the character following When parentheses are used to group a pattern, recalls items stored in pattern memory</td>
<td>: A literal dot \: A literal backslash \1: Recalls 1st pattern stored in parentheses</td>
</tr>
</tbody>
</table>

Table F.2 Metacharacters that can be used with the GenomBench Pattern Search Tool
<table>
<thead>
<tr>
<th>Metacharacter</th>
<th>Function</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ ]</td>
<td>Matches any one of the enclosed characters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- A hyphen (-) indicates a range of consecutive characters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- A caret (^) as the 1st character in brackets matches any one character not in the list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- A hyphen (-) or right bracket (]) as the 1st character is treated as a member of the list</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All other metacharacters are treated as members of the list</td>
<td></td>
</tr>
<tr>
<td>[0-9]</td>
<td>Any digit</td>
<td></td>
</tr>
<tr>
<td>[^a-z]</td>
<td>Any character except a lower case letter</td>
<td></td>
</tr>
<tr>
<td>[^0-9-]</td>
<td>Any character EXCEPT a digit, hyphen, or bracket</td>
<td></td>
</tr>
<tr>
<td>*</td>
<td>Matches zero or more of the single characters immediately preceding it. Can also match a regular expression preceding it.</td>
<td></td>
</tr>
<tr>
<td>A*</td>
<td>Zero or more upper case As.</td>
<td></td>
</tr>
<tr>
<td>[AAA]*</td>
<td>Zero or more strings of AAA</td>
<td></td>
</tr>
<tr>
<td>+</td>
<td>Matches one or more instances of the previous regular expression</td>
<td></td>
</tr>
<tr>
<td>A+</td>
<td>One or more upper case As.</td>
<td></td>
</tr>
<tr>
<td>[A]+</td>
<td>One or more occurrences of A</td>
<td></td>
</tr>
<tr>
<td>?</td>
<td>Matches zero or one instances of the preceding regular expression</td>
<td></td>
</tr>
<tr>
<td>A?</td>
<td>Zero or one upper case As.</td>
<td></td>
</tr>
<tr>
<td>[A] +</td>
<td>Zero or one occurrences of A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Matches the regular expression before or after</td>
</tr>
<tr>
<td>CAGT</td>
<td>ACGT</td>
<td>The string CAGT or the string ACGT</td>
</tr>
<tr>
<td>()</td>
<td>Applies a match to the enclosed group of regular expressions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stores an expression into pattern memory for later retrieval</td>
<td></td>
</tr>
<tr>
<td>([a-z]+)</td>
<td>One or more instances of any lower case character</td>
<td></td>
</tr>
</tbody>
</table>

*Table F.2 Metacharacters that can be used with the GenomBench Pattern Search Tool (continued)*
Once you have installed Vector NTI Advance 10, you will need to license the application to be able to use it.

To satisfy the needs of users in different industrial, scientific or educational environments, Invitrogen has designed three types of Vector NTI licenses, as well as a Demo Mode option. These are all administered through the License Manager.

- **Static License**: Purchased by one user for installation on one computer
- **Dynamic License (DLS)**: A license that is installed on a server and issued by that server to client Vector Advance computers. DLS licenses are shared by a specified number of users or “seats,” with the number of users at any one time being limited to the number of “licenses” specified in the contract.
- **Trial License**: Allotted to a potential purchaser of Vector NTI software for a specified number of days, during which the user can review and use the software within certain limits.
- **Demo Mode**: When unlicensed, Vector applications operate in Demo Mode for the purpose of demonstrating the basic features of the software. Some functionality is disabled in Demo Mode.

When you open the Vector NTI Advance software, a checkmark icon, such as this (✓), at the bottom right corner on the Status Bar shows the current license status. Pause the cursor arrow over the button, and a pop-up label displays the license status.

- Green checkmark = active Static License
- Green, *blinking* checkmark = active Trial or Dynamic License
- Red, *blinking* X = the application is not licensed; running in Demo mode.

License Manager does not open automatically when you install Vector Advance on your computer. You must open License Manager manually. To open License Manager, select it from the Start menu: **Start > Programs > Invitrogen > Vector NTI Advance 10> License Manager** or click **Help > License** in the Database Explorer or a Vector NTI viewer window once you have opened it.
License Manager Dialog Box

The License Manager has three tabs, the **Contact Us** tab, the **Personal** tab, and the **Applications** tab.

**Contact Us Tab**

The License Manager opens by default to the **Contact Us** tab (Figure G.1). This tab summarizes your Vector software licensing agreement. Additionally, it provides information for upgrading your Vector application license and contacting Invitrogen.

![Figure G.1 License Manager (Contact Us tab)](image)

**Personal Tab**

The **Personal** tab (Figure G.2) provides text boxes for entering personal information. Once entered on this tab, when you click your license choice on the **Applications** tab, your entries are automatically entered on the license application.

![Figure G.2 License Manager (Personal tab)](image)

**Applications Tab**

The **Applications** tab (Figure G.3) indicates the type of License currently in effect for each Vector NTI Advance application, as well as for Vector Xpression and PathBlazer. If you have Vector
For a new installation or update of a previously unlicensed installation, License Manager opens in Demo mode for all applications.

For Static, Dynamic and Trial licenses, if you are not licensing the entire software package using the same type of license, click in the license-type text box of the application for which you wish to specify a license. Click the down-arrow to extend the drop-down menu and select the appropriate license type. (Demo mode is the default selection.)

**Note:** If your active license type varies for different applications, this dialog box may have a combination of rows with drop-down menus and rows displaying contract numbers. Click the button appropriate for the license type you want to register. Each option is described in the following sections.
Static License Dialog Box

Static licenses can be obtained in either a personal or a corporate mode. To configure a static license, click the Static button at the bottom of the Applications tab (Figure G.3). This opens the Static License dialog box (Figure G.4).

For either type of static license, enter your name, organization, phone number and email address in the appropriate fields. This sets the user information in Vector NTI Advance.

Note: If you already entered your personal information on the Personal tab, it should appear here when you open this dialog box.

In the License # field, enter your Vector NTI Advance 10 static license number provided in the e-mail you received if you downloaded the software from the Invitrogen website, or in the letter that accompanied your CD ROM and/or manual.

- **Personal Static License**: You must have an internet connection for this step. Click the Apply button. The registration Key is returned by the TCP/IP protocol right after you press the Apply button, and it will be applied automatically.

  If the registration fails because of a missing connection to the Invitrogen licensing server, an appropriate message immediately displays. In such a case, you can contact Invitrogen Customer Support or Sales, providing them your computer’s hardware ID and your license number.

  Once you receive the registration key, enter the key in the Key text box of the Static License dialog box. Make sure the License Number is entered appropriately, and click Apply. If the Key matches your license number and computer hardware ID, the license is registered.

- **Corporate Static License**: No internet connection is needed nor is a Registration Key required. The License Manager on your client computer validates the Corporate License number; if it is valid, then your software is registered.

  Once you receive the registration key, enter the key in the Key text box of the Static License dialog box. Make sure the License Number is entered appropriately, and click Apply. If the Key matches your license number and computer hardware ID, the license is registered. No connection to the Internet is required in this case.

Notes:
- Once you have applied your static license, notice that the Applications tab reflects your static license status.
If you want to reset your personal static license, (you cannot re-set a corporate static license) type *Unregister* in the License Number field and click *Apply*. You will be warned that you are trying to reset your static license and asked if you want to continue. If you answer *Yes*, the application will reset your license and will send proof of this operation to the Invitrogen server. If the connection to the server fails, you will receive notice of this.

**Dynamic License Dialog Box**

Technically, a Dynamic License is not a "license" mode, but registers permission to work in a shared DLS environment in license mode. To configure your Dynamic license, click the **Dynamic** button at the bottom of the **Applications** tab (Figure G.3). This opens the Dynamic License dialog box (Figure G.5):

![Dynamic License dialog box](image)

1. Enter your name, organization, phone number and email address in the appropriate fields. This sets the user information in Vector NTI Advance 10.

**Note:** If you already entered your personal information on the **Personal** tab, it should appear here when you open this dialog box.

2. In the **URL of DLS** text box, enter the DLS server URL supplied by the DLS administrator at your site. If your DLS server requires a password, make sure the authentication settings are filled in appropriately.

3. Press the **Internet Connection Settings** button to configure your connection settings and to enter server proxy information, if a firewall is used at your site. Refer to the following section, **Internet Connection Settings**.

4. Press the **Test Connection** button to review the status of your connection. Refer to **Testing the License Server Connection** on this page.

5. Once you have configured the **Dynamic License** dialog box parameters, to set all Vector applications to Dynamic License, press the **Set For All Applications** button. Once you do this, when you close this dialog box, the Applications tab now shows Dynamic License for all applications.

**Note:** When you set Dynamic licenses for all applications, this operation only applies for those applications for which you do not have a Static License.

Press the **Apply** button to execute the dynamic license configuration.
Internet Connection Settings

For a Dynamic License, press the Internet Connection Settings button in the Dynamic License Server Connection dialog box. This opens the Internet Settings dialog box where you can alter your proxy settings (Figure G.6):

![Internet Settings dialog box](image)

The Internet Settings dialog box allows you to set your connection parameters. If the Use Internet Explorer settings button is selected, License Manager will attempt to make the connection using your default settings. If default detection is not successful, you can either choose the Direct connection button if you do not have a proxy or choose the Use proxy server button and specify the proxy name, port and password information.

Press the OK button to return to the Dynamic License Server Connection Tester dialog box. See the Internet Connection Settings section on page 692 for more information.

Testing the License Server Connection

In the Dynamic License dialog box, press the Test Connection button to review the status of your connection. This opens the Server Connection Tester dialog box (Figure G.7):

![Server Connection Tester dialog box](image)

The status of the connection displays in the right-hand panel.

- If the server requires a password, it must be entered into the corresponding text box in this dialog box.
- If you want to alter your proxy settings, press the Internet Connection Settings button (see next section).
- Once the settings are reconfigured, press the Connect button to test the connection using the new settings.

Once you have configured the Dynamic License dialog box parameters, to set all Vector applications to Dynamic License, press the Set For All Applications button. Once you do this, when you close this dialog box, the Applications tab now shows Dynamic License for all applications.
**Note:** When you set Dynamic licenses for all applications, this operation only applies for those applications for which you do not have a Static License.

Press the **Apply** button to execute the dynamic license configuration.

**Trial License Dialog Box**

To configure a trial license, click the **Trial** button at the bottom of the **Applications** tab of License Manager (Figure G.3). This opens the Trial License dialog box (Figure G.8):

![Trial License dialog box in License Manager](image)

As described in the Trial License dialog box, you must generate an e-mail addressed to `bioinfo-trial@invitrogen.com` and include your name, the name of your organization, phone number, and email address.

**Notes:**

- The e-mail will not be generated automatically from this dialog box.
- You will receive a trial license number in response from an Invitrogen contact person. Return to this dialog box and enter the trial license number in the text box. Click the **Apply** button to complete the trial license registration.
- When you set Trial Licenses for all applications, this operation only applies for those applications for which you do not have a Static License.
Numerics

3D-Mol
- 3D Options dialog box 616
- accepted file formats 605
- Align Atoms mode 621, 622
- Align Backbone mode 621
- aligning molecules in 620
- alignment viewing options 621
- alternating between panes 606
- atoms display 611
- backbone display 611
- calculating protein surfaces 622
- color scheme display 612
- color scheme, creating custom 612
- copying chains 608
- creating user-defined structures 607
- deleting structures 608
- display window 606
- Editable Entry 607
- elements of 606
- exporting structures 624
- Graphics Pane 610
- hide options 616
- manipulating graphics 611
- mark types 615
- marking signals 609
- marks display 615
- measuring graphics 614
- molecules view 614
- moving molecule 610
- NewEntries 607
- opening structures in 606
- painting/unpainting marked signals 609
- Project Pane 607
- removing panes from view 607
- resizing molecules 610
- saving results 624
- sending molecules to VNTI 623
- Sequence Pane 611
- setting mark radius 610
- show options 616
- stereo viewing option 620

Structure Pane 609
- surface calculation references 661
- surface display 612
- tools 620
- viewing file properties 607
- viewing graphics 611

A
- abi files 132
- Absolute Complexity
  - graphs in AlignX 393
  - statistical significance 393
- Activating
  - window panes 15
- Adaptor fragments 315
- Add to Oligo List
  - Order custom primers 72
- Adding
  - features to molecule 236
  - gel marker 369
  - gel sample 367
  - molecules to alignment 401
  - molecules to an alignment 403
  - molecules to project, AlignX Blocks 598
  - new oligo 285
  - restriction site 242
  - restriction sites to cloning fragments 314
  - restriction sites to molecule 236
  - separators, Database Explorer 167
  - to Oligo List 283

Advanced analyses
- BLOCKS 565
- copying 570
- deleting 570
- editing 570
- operations 570
- Pfam 565
- ProSite 568
- Proteolytic Cleavage 566
- reading 570
- saving 571
- sending to GenomBench 571
sending to Vector NTI 571
Sim4 562
Spidey 564
starting an analysis 570
stopping an analysis 570
viewing in GenomBench 572
viewing in Vector NTI 571

Algorithms
alignment 661
phylogenetic tree 662
protein analyses, BioAnnotator 427

Alignment
3D-Mol 620
adding molecules to 403
alignment PCR 260
basic algorithms 661
changing sequence order 399
color key 400
displaying feature alignments 523
displaying Sim4/Spidey hits in GenomBench 542, 544
editing 408
editing, tutorial 120
finding features in 391
finding sequences in 398
finding signals in 392
graph in AlignX 393
importing, MSF format 402
initiating multiple 402
initiating profile 403
linking panes 396
marking features in 391, 396
matrix, tutorial 118
multiple, parameters 405
pairwise parameters 404
performing on Internet, tutorial 110
protein profile, parameters 406
removing a molecule 403
score 646
score, tutorial 118
setting parameters 403
template files 409
viewing options, 3D-Mol 621

Alignment Display Setup
color setup 401
Identity Tables 400

Alignment Pane
AlignX 396
AlignX Blocks 597
AlignX Blocks, selection in 598
CE, description 480
CE, editing fragment or consensus 483
Contig Window 479
copying data 20
linked panes, AlignX 396

Alignment PCR
description 260
launching from AlignX 261
Lists dialog box 261
process 260
report 263

Alignment projects
adding molecules to 401
creating new 401
opening 401

AlignX
adding graphic analyses 117
adding molecule to alignment 118
algorithms 402, 413
alignment colors 119
Alignment PCR 413
Alignment Setup 118
alternating between panes 390
Analysis Pane 393
changing sequence order in Alignment Pane 399
consensus calculation 399
creating custom analyses 394
designing PCR primers 413
Dot Matrix calculation 410
editing alignment, tutorial 120
editing an alignment 408
elements of display window 390
Feature Type Color Scheme dialog box 397
finding sequences 398
Graphics Pane 393
graphs 393
guide tree 392, 406
managing scales of graphs 395
marking features in 391, 396
opening 389
opening from DB Explorer, tutorial 115
Plot Setup 394
realigning selected area 403
reverse complementing sequences 398
saving projects 122
scoring matrices 407
identity table, pairwise comparison 409
Identity Tables 400
tutorial 115

AlignX Blocks
adding molecules to project 598
Alignment Pane 597
alternating between panes 594
block color setup 601
Block List Pane 595
Block Search 598
Block Sequence Space 599
color setup for links 599
defining block manually 599
deleting a molecule 601
description 593
elements of display window 593
finding Text Pane features 595
linking/unlinking residues 599
opening a project 598
removing blocks 596
Schematic Pane 597
Score Matrix 602
selection techniques 596
setup parameters 601
sorting blocks 596
templates 602
Text Pane 594
Amino acids See Proteins 8
Analyses
creating custom, BioAnnotator 428
description of, BioAnnotator 426
DNA/RNA, BioAnnotator 426
in reports 377
protein, BioAnnotator 426
ProtScale 427
Analyses Setup 221
Analysis
adding or removing, BioAnnotator 423
Back Translation 297
changing properties, BioAnnotator 423
Find Common Non-Cutter REN dialog box 302
Find Common Non-Cutter REN report 304
loading PCR analysis settings 247, 328
Non-Cutter REN dialog box 304
RFLP 300
saving PCR analysis settings 247
Analysis List setup
BioAnnotator 423
Analysis Monitor
Analysis List Pane 562
analysis operations 570
BLAST Search 569
BLOCKS parameters 565
configuring parameters 562
copying analyses 570
deleting analyses 570
description 559
editing analyses 570
launching 560
launching from GenomBench 542
Parameters Pane 561
Pfam parameters 565
ProSite parameters 568
Proteolytic Cleavage parameters 566
reading analyses 570
saving analyses 571
sending to GenomBench 571
sending to Vector NTI 571
Sim4 and Spidey alignment analyses, GenomBench 542
Sim4 parameters 562
Spidey parameters 564
starting an analysis 570
stopping an analysis 570
updating search databases 559
viewing in GenomBench 572
viewing in Vector NTI 571
window description 561
Analysis Pane
BLAST Viewer 585
changing parameters, BLAST Viewer 587
selecting regions, BLAST Viewer 587
Analysis Results
Database Search 178
Analyzing
hybridization probes 280
PCR primers 247
sequencing primers 276
Annotations
adding 50, 227
deleting 227
editing 227
Annotations folder 208
API
description 5, 11
manual on website 11
AppleTalk
shared database 383
Archive
data stored in 164
importing enzymes from 183
Arrows
meaning of direction/DNA graphics 208
ASCII Format 667
assembly results 134
Assembly Setup
ContigExpress 444
Light Settings tab 450
Assembly Setup dialog 134
Assembling contigs 442
Authors
foreign 170
local objects 169
modifying 169, 170, 186
object system field 169
Automation, APIs 5, 11
B
Back Translation
changing codon usage tables 299
codon ambiguity 299
codon usage tables 298
copy options 299
creating codon usage tables 298
description 297
dialog box 297
697
opening files in 297
repairing codon usage tables 299
translation tables 299
tutorial 122
Basic molecules 305
Bibliography
Copy Bibliography 166
Copy Tag 166
Bibliography conversion 165
BioAnnotator
adding analyses, tutorial 129
advanced analyses 559
alternating between panes 416
analyses descriptions 426
analysis list setup 423
Analysis Monitor 559
Analysis Monitor analyses 428
BLOCKS parameters 565
BLOCKS search 428
changing analysis properties 424
copying sequence to clipboard 422
creating custom analyses 428
creating project 422
data selection 419
description 415
editing properties 419
features of 416
formatting sequence, tutorial 128
launching 415
modes in Plot Pane 417
modifying layout, tutorial 128
modifying properties, tutorial 127
opening a project, tutorial 129
opening existing project 422
opening from VectorNTI 125
overlapping plots 418
overlapping plots, tutorial 128
panes in 416
Pfam parameters 565
Pfam search 428
Plot Pane 417
ProSite parameters 568
ProSite search 428
protein analyses algorithm 427
protein analyses setup 424, 425
proteolytic cleavage analysis 428
Proteolytic Cleavage parameters 566
sequence display setup 421
Sequence Pane 420
tutorial 125
Biochemical operations
fragment termini 318
BLAST
Internet search, tutorial 107
reference 665
search tutorial 108
Ultimate ORF BLAST 290
Viewer 583
Blast
Ultimate ORF 290, 467
BLAST Results
Database Search 178
BLAST Search
Analysis Monitor 569
Analysis Monitor launch 569
BLAST 2.0 576
description 569, 575
dialog box 576
expectation value 579
filter masks 579
GenBank division identifiers 577
matrices 580
MegaBLAST 577, 581
Parameters tab 578
performance 582
program descriptions 577
PSI-PHI BLAST 576
PSI-PHI tab 581
saving results 589
saving results, tutorial 110
Sequence tab 578
types description 576
viewing options 582
web site reference 569, 575
word size 579
BLAST Viewer
Alignment Pane 587
Analysis Pane 585
Color Setup 588
consensus calculation 588
copying molecules to the clipboard 591
creating new subset 591
deleting hits 585
displaying hits as features 590
elements of 583
exporting data from 591
exporting hits as features 590
exporting to GFF file 590
graphs 585
Hit Distribution graph 586
hit element description 585
Hit Map graph 586
hit molecule information 585
Identity Table 588
launching tools from 591
opening selected hits 590
Plot Setup 587
print command 591
Query-Hit Alignment graph 586
saving selected hits 590
Select Hits feature 587
selecting Analysis Pane regions 587
selection techniques 588
Sequence Profile graph 586
sorting hits 584
Text Pane 584
zooming graphics 586
Block
defining in AlignX Blocks 599
in AlignX Blocks, description 593
Block Editor
AlignX Blocks 600
Block Search
AlignX Blocks 598
Block Sequence Space
in AlignX Blocks 599
BLOCKS
parameters 565
updating BLOCKS database 559
Blocks List Pane
AlignX Blocks 595
description 596
selection techniques 596
BLOSUM 646, 662
Broadcast Selection
accessing 19
feature 19
for sequence pane selections 19
Buffers
Gel Display 366

C
Cache
user field 198
Calling Secondary Peaks
chromatogram 501
executing changes 503
List Pane 501
Sequence Pane 502
Camera
copying with 19
Caption
Figure 17.16 Citations Database Search dialog box 178
Figure 17.17 BLAST Database Search dialog box 178
Figure 17.9 Ancestors Setup dialog box 174
Caret
in Graphics Pane 235
in Sequence Pane 236
caret position 144, 147
Changing
Display Setup 216
Graphics Pane formatting 226
Graphics Pane settings 211
sequence text 215
starting coordinate 244
starting coordinate, tutorial 42
symbol sizes 228
Chromatogram
changing scale of 470
editing a sequence 470
editing summary 470
evaluating quality 478
show/hide 482
chromatogram data 141
Chromatogram Pane 141
Fragment Window 469
navigating 469
chromatogram peak heights 143
chromatogram traces 143
Circular Display 208
Citation
converting journal format 166
creating new 190
Database Search 178
Citation Viewer 634
Clearing
subset 181
User Fields 200
Cleavage points, entering 187
Cloning
component fragments 313
component fragments in 356
create all possible constructs 318
Dummy fragments 313
General Information 312, 356
linker fragments 314
sequence fragments 314
Cloning See also Construction or Design 349
Closing
Text Pane folders 208
Codon ambiguity
Back Translation 299
Codon usage
Back Translation 298
changing codon usage table 299
creating custom table 298
repairing corrupt tables 299
translated molecules 294
Collapsing folders 16
Color key
alignment 400
Color setup
AlignX Blocks 601
BLAST Viewer 588
for links in AlignX Blocks 599
Colors
creating custom scheme, 3D-Mol 612
defining custom colors, 3D-Mol 614
Columns
modifying in CE Project Explorer 457
modifying/Open/Save As dialog boxes 195
Contig Graph 151
Coverage Map 151
creating assembly 151
deleting data 464
deleting fragments 140
demo project 132
editing fragment or contig 484
exiting 140
exiting a project 460
exporting contigs/fragments from 460
features of workspace windows 461
finding ambiguous bases 481
finding sequence fragments 467
finding sequence fragments in Alignment Pane 481
finding Text Pane feature in other panes 463
fragment trimming 493
Fragment Viewer 140
Graphics Pane 151
highlighting a fragment 485
History View 139
importing a directory of files 441
importing fragments 132, 133, 438
importing Phrap file 439
importing Phred file 439
launching, tutorial 131
List Pane 133, 137
modifying viewing options 136
opening a project 438
opening fragment and contig windows 461
Options dialog 136
Project Explorer 133, 134
Project Explorer, List Pane 132
Project window options 457
references 665
saving a project 140
saving items in project 464
saving projects in 131
selecting objects in 442
showing ORFs, Contig window 478
showing ORFs, Fragment window 467
showing two strands 142
starting 131
Status Bars 462, 584
translating consensus, Contig window 485
tutorial 131
Ultimate ORF Blast 467
viewing options 136, 153
Weight Graph 151, 152
zooming in 151
ContigExpress projects
opening 132
ContigExpress window 131
Converting GCG files to VNTI 643
Copying analyses, Analysis Monitor 570
Back Translation 299

bibliography format 166
chains in 3D-Mol 608
Copy To command 20
database data 191
formats for 20
fragment/contig in Tree Pane 459
gel data 373
panes and text from GenomBench 557
raw sequence 20
sequence 215
sequence in GenomBench 522
styles 230
to clipboard 19
using Camera 19
with Copy button 20

Creating composite features, GenomBench 520
Construction fragments 316
custom genetic code 295
features, GenomBench 517
Gel Display window, tutorial 99
gel sample 367
internal start point 212
molecule shortcuts 232
new Citation 190
new DNA/RNA molecule 186
new enzyme 186
new enzyme, methylation 187
new fragment 316
new Gel Marker 190
new molecule from sequence 237
new molecule in Design 349
new molecules 7, 306
new oligo 189
new protein molecule 186
new subset 179
project in AlignX 401
project in AlignX Blocks 598
project, BioAnnotator 422
user-defined structures, 3D-Mol 607
Creating a New Destination Vector 335
Creating Entry Clones 323
CSV Oligo Files Importer 648
Custom Primers
Invitrogen 65, 72
Customer support, contacting 659
Cut/copy sequence 215

D
DAS Configuration utility 507
DAS servers 507
Data Loading Monitor, GenomBench 511
Database
author information 169
backup 179

701
Vector NTI Advance 10 User’s Manual

capability 5
cleaning up 179
Contact Manager 63
copying data 191
cut, copy 191
description, objects 167
drag and drop feature 191
LabShare, configuring link to VNTI 637
LabShare, viewing in VNTI 639
management operations 178
menu commands for objects 185
missing Gateway Cloning molecules 322
moving objects 191
object comments 170
object keywords 171
object system fields 168
object user-defined fields 170
operations on objects 185
pasting data 191
Proteins 56
restore 179
search 6, 171
search, tutorial 58
Subsets 56, 58
types of molecules 165
User Field Manager 59
Database Explorer 57
 Archives tab 193
column display 192
Contact Manager 195
content reports 193
description of 163
elements of 165
exchanging objects from 164
Gateway Cloning launch 186
modifying window display, VNTI 192
molecule export preferences 193
opening molecule from 205
popup tips display 193
searching in 171
shortcuts 194
sorting columns 193
Status Bar display 193
Subsets 167
tables 165
User Field Manager 197
viewing modes 192
Database Search
 Analysis Results 178
Ancestors Setup, VNTI 174
Attributes Setup, VNTI 173
BLAST Results 178
citation search 178
defining filters, VNTI 172
DNA/RNA 172
enzymes 176
Features Setup, VNTI 175
Gel Markers 177
Keywords Setup, VNTI 174
name search, VNTI 172
oligos, VNTI 172
protein 172
steps in, VNTI 171
DatabaseSee also Shared database 386
Date/time display
editing 170
modifying settings 186
DB Explorer
opening at startup 13, 26, 56, 164, 204, 234
deleted peaks 147
Deleting
 analyses, Analysis Monitor 570
custom genetic code 296
data from ContigExpress 464
features, GenomBench 517
gel lane 372
hits, BLAST Viewer 585
molecule in AlignX Blocks 601
molecule with descendants 186
object from database 185
sequence regions 215
subset contents 181
tool 626
Demo Mode 650, 687
demo project 132
Design
 Advanced Preferences 360, 362
advanced, tutorial 93
biochemical operations 353
complex donor, tutorial 97
complex recipient, tutorial 93
component fragments 356
defining 3’ end 352
defining 5’ end 351
defining component fragments 356
defining donor fragments 355
defining fragments 235
defining fragments, tutorial 86
defining recipient fragment 350
describing goal molecule 350
description folder, tutorial 90
Design Description folder 362
Donor 309
Flank regions 357
flanking regions 354
fragment descriptors 357
Fragment Editor 353
fragment isolation 359
alignments, GenomBench 525
analyses, Analysis Monitor 570
analysis properties, BioAnnotator 419
Analysis Result properties 190
BLAST Result properties 190
blocks in AlignX Blocks 600
changing Styles 51
chromatogram sequence 470
chromatogram, summary 470
contact information 196
contig project properties 452
contig properties 452
date/time display options 170
display properties, BioAnnotator 419
Feature map 240
Feature map, tutorial 41
feature types 49
features, GenomBench 517
fragment or contig properties 463
fragments or consensus, CE 483
fragments, CE 483
gel properties 371
graphics 48
Graphics Display settings 222
molecule editing vs picture editing 233
molecule feature 236, 240
molecule fields, tutorial 38
molecule in Molecule Editor 233, 237
molecule, effect on parents 238
molecule, tutorial 37
Molecules 7
motifs 217
object comments 170
oligos 189, 217, 285
options, ContigExpress 457
options, Vector NTI 204
options, VNTI DB Explorer 192
pane layout 22
Picture Editing Mode 48
restriction map 242
reversing molecule edits 236
sequence 236, 240
sequence attributes, GenomBench 528
sequence, effect on feature 40
sequence, tutorial 39
shared db properties 386
shared db user accounts 386
styles 230
subset 180
subsets, Resulting Content column 180
subsets, Source Content column 180
subsets, Source Subsets column 180
symbol format 228
User Fields 200
Electrophoresis profile 366

Enlarging
  graphics selection 18
Ensembl
  extracting features in GenomBench 513
Entrez/PubMed Search 631
Entry Clones
  pDONR vectors 323
Enzyme
  creating new enzymes 186
  Database Search 176
  DB operations on objects 185
  entering cleavage points 187
  entering methylation points 187
  entering Recognition String 187
  methylase 187
Enzyme Editor 186
Exons
  graphic display 210
  splicing into new molecule 307
Expanding folders 16
Expectation value 579
Exporting
  format 193
  fragment/contigs from CE 460
  hits as features, BLAST Viewer 590
  into text file directory 184
  molecule 232
  subset 184
  subset to VNTI Archives 184
  to GFF file, BLAST Viewer 590
User Field Definitions 199
User Fields 60
Expression Clones
  Recombine by BP 323
Extra-Chromosome Replication 313, 356

F

FASTA files 442
Feature map
  adding new feature, tutorial 42
  data in reports 378
  editing 240
  editing, tutorial 41
  folder 207
  sequence editing effects 40
Feature Map Setup
  DNA/RNA 216
  proteins 223
Features
  adding to an alignment, GenomBench 520
  assigning Styles 230
  downloading from NCBI, GenomBench 520
  editing 48, 240
  editing graphics 229
  extracting component features, GenomBench 513
marking in AlignX Alignment Pane 391, 396
searching by in VNTI Database 175
transforming to permanent, GenomBench 543
translating 293, 294
File formats 204
accepted in ContigExpress 439
Files
acceptable formats, AlignX 402
exporting, BioAnnotator 22
importing, AlignX 402
Find
Common Non-Cutter REN Analysis 302
Finding
ambiguous bases, CE 481
molecule features 208
objects/Find button 17
sequence fragments in Alignment Pane, CE 481
sequence, DNA 214
sequence, protein 214
sequences, AlignX 398
shared database 385
specific sequence, Sequence Pane 467
Text Pane feature in other panes, CE 463
Text Pane features, AlignX Blocks 595
Fit to Window
Graphics Pane 18
Folders
Hybridization Probes 281
motifs search 291
mutagenesis 292
opening/closing 16
ORFs results 290
PCR results 271
restriction fragments 300
Text Pane 207
Text Pane, AlignX 391
Font
Editing 47
Footer
Codes for 21
Formats
ASCII 667
Bairoch (enzymes) 182
Database Explorer reports 194
feature table descriptions 182
files 204
imported text files 182
IUB codes 667
text file directories 182
tools input 627
VNTI Archive 182
Formatting
bibliography 166
Feature Types 50
graphical features 226
Graphics Pane text 227
manuscript citations 166
sequence text 18
Formatting Codes
Header and Footer 21
Fragment
editing properties of, CE 463
importing directory to ContigExpress 441
importing to ContigExpress 438
importing to ContigExpress, accepted formats 439
Fragment Descriptor
Dialog Box 357
Fragment Editor
in Design 353
Fragment of Molecule
dialog box 316
Fragment trimming
adding insertion points 499
calculating preview 497
confirming with chromatogram 498
description 493
editing polylinker bases 499
initiating in Project Explorer 456
settings 495
steps in process 493
vector contamination 498
Fragment Viewer
Chromatogram Pane 157, 469
deleted peaks 147
description, CE 465
displaying ORFs 144
editing sequence 147, 149, 159, 160, 161, 470
finding sequence fragments 467
modifying viewing options 472
nucleotide sequences 145
opening 140, 461, 465
Sequence Pane 466
setting caret position 144, 147
showing individual traces 143, 144
showing two strands 144
showing deleted peaks 471
showing ORFs 467
Text Pane 466
Fragment Wizard
defining 3' end, Construction 310
defining 5' end, Construction 310
defining fragments 309
in Construction 309
in Design 350
Insert Inverted box 309
New Fragment description 311
selecting restriction site 352
steps in using 309
steps in using, Design 349
Fragments
adding new, Construction 313
Call Secondary Peaks 456
Functional features
   Sequence Pane display 211
   Functional Signal Name Dialog Box 241

**G**
Gap penalties 405
Gateway Cloning
   creating Expression Clones 323
   launching from DB Explorer 186
   missing molecules 322
   Recombine an Expression Clone by BP 323
   Recombine Expression Clones by BP 323
GCG Converter 643
   importing converted files 645
Gel
   changing fragment color, tutorial 102
   copying 373
   creating marker, tutorial 101
   gray zone around fragments 373
   partial digestions 368
   Text Pane folders, description 102
Gel Display
   electrophoresis parameters 366
   electrophoresis profile 366
   electrophoresis type 366
   lanes in 372
   parameters 366
   selecting fragments 372
   True Scale 372
Gel Display Setup 366
Gel Display window 8
   calculator 373
   camera feature 373
   copying data 373
   creating new 8
   elements of 367, 370
Gel Pane 370
   opening 365
   printing in 374
   Text Pane 370
   toolbars 367
   tutorial 99
Gel document
   opening 367
Gel electrophoresis
   adding samples, tutorial 100
   animation 371
   animation, tutorial 103
   calculator, tutorial 105
   changing color of fragment, tutorial 102
   fragment separation time 372
   manipulating lanes 372
   markers 8, 367
   modifying fragment properties 371
   opening a new gel 366
running 9, 371
running, tutorial 103
samples 8
separation distance 373
separation distance, tutorial 104
separation time 373
separation time, tutorial 104
setup, tutorial 99
time 371
Zoom in/out 372
Gel marker
   adding 369
   creating 368
   creating new 190
   Database Search 177
description 367
   editing 190, 369
   New Gel Marker dialog box 368
   saving 368
Gel Marker Editor 370
Gel Pane
   Gel Display window 370
Gel sample
   adding 367
description 367
   list 369
   performing partial digestions 368
Gel Sample List
   adding fragments to 370, 372
   adding fragments to gel 370
   viewing 369
GenBank
   division identifiers 577
Generate a Clone by PCR Amplification and TOPO Cloning 339
Generate a Clone by TOPO cloning 337
Generate an Entry Clone by PCR Amplification and BP Recombination 324
Generate an Entry Clone by PCR Amplification and Ligation into a pENTR Vector 329
Generate an Entry Clone in a BP reaction by Recombining an Expression Clone 323
Genetic Code
   alternate code list 295
   changing 295
   creating custom 295
deleting custom 296
dialog box 295
   naming custom 295
   selecting 175
   setting for translations 214, 294
GenomBench
   activating panes 513
   advanced analyses 559
   Analysis Monitor 559
configuring DAS server paths 508
copying from 557
creating local databases 506
creating new sequences 552
DAS Configuration utility 507
Data Loading Monitor 511
database file 506
database migration 506, 554
editing fragment loading parameters 534
editing sequence attributes 528
exporting fragments 556
extracting Ensembl features 513
feature type filters 530
highlighting sequence regions 514
importing files to local database 554
launching 506
launching Analysis Monitor 542
launching Query Application 545
loading features from Ensembl 513
loading fragments and features 510
LS Configuration utility 506
printing from 556
redo operations 529
removing and restoring panes 513
saving fragments to local database 552
selecting global features for download 536
sequence on demand mode 535
setting selections and fragments 514
Sim4 parameters 562
Spidey parameters 564
Spiedy parameters 564
target sequences 521
undo operations 529
zooming panes 514
GenomBench Alignment tab 523
adding features to alignments 520
alignment color scheme 524
description 523
editing alignments 525
GenomBench Feature Map Pane
adding features to Alignment tab 523
adding files and URLs to features 518, 521
adding qualifiers to features 519
composite features 520
coordinate scale options 517
creating features 517
deleting features 517
description 515
editing features 517
editing graphic styles 531
editing multiple feature simultaneously 520
feature layout options 516
feature style options 517
grouping features 520
marking selections 514
GenomBench Information Pane
adding features to Alignment tab 523
description 521
Features tab 526
Search Results tab 540
Sequence tab 522
Text tab 528
GenomBench Overview Pane
description 515
GenomBench Search Results tab
description 540
GenomBench Sequence tab
copying sequence 522
description 522
translating sequence 523
GenomBench Text tab
description 528
GenomBench tools
Feature Search 539
Pattern Search 540
regular expression syntax, Pattern Search tool 683
retrieving Sim44/Spidey alignment analysis results 542
sending fragments to Vector NTI 551
showing Sim4/Spidey hits in Alignment tab 544
transforming features to permanent 543
Global alignments 646
Goal Molecule Definition List 355
opening 311
Graphical map
fit to window 18
Graphics
assigning styles 230
formatting 226
manipulating in 3D-Mol 611
modifying properties, BLAST Viewer 587
modifying styles 230
Graphics display
editing 222
settings 222
Graphics objects
formatting 48
Graphics Pane
3D Options dialog box, 3D-Mol 616
3D-Mol 610
AlignX 393
CE, resizing graphics 476

specifying URLs for features with alignment information 537
unmarking selections 514
viewing feature properties 520
changing settings 211
Contig Window 474
editing modes 226, 233
finding objects 17
formatting features 226
hiding features 229
Molecule Display window 208
moving objects 228
re-sizing 17
restoring features 229
selection techniques 234
styles 229

Graphs
AlignX 393
BLAST Viewer 586
descriptions 393
Hit Distribution 586
Hit Map 586
managing scales of, AlignX 395
Query-Hit Alignment 586
Sequence Profile 586

Guide tree
algorithm 662
AlignX 392
construction options 406
exporting 392
reason for empty pane 392

Guide Tree Pane
tutorial 118

Hairpin loops
description 288

Header
Codes for 21
Header and footer codes
page setup 21
High throughput cloning 347
High Throughput Operations 344
High Throughput PCR 344
Hit Distribution Graph 586
Hit elements
BLAST 585
Hit Map Graph 586
HTML publishing 232
Hybridization probe
user-defined 280
Hybridization probes
analysis 279
dialog box 246, 279
folder 281
parameters 280
Hybridization probes folder 207

Identity Table
AlignX 400, 409

Importing
ASCII 667
directory of fragments into ContigExpress 441
enzymes from REBASE 182
enzymes from VNTI Archive 183
features 182
file formats 182
fragments into ContigExpress 438
molecule files 181
Molecules 7
molecules to subset 182
objects from text file directories 182
objects from VNTI Archives 182
oligo files, CSV Importer 648
Phrap file into ContigExpress 439
Phred file into ContigExpress 439
Phred quality values into ContigExpress 441
REBASE commercial source data 183
User Field Definitions 199
User Fields 60

Inserting
sequence 215, 236

Internal Design Parameters, Design 360
Internal start point
creating in DNA/RNA Sequence 212

Internet 52
saving results in database, tutorial 112

Internet Tools 10

Introduction to Gateway Cloning 321, 322
Introduction to TOPO Cloning 337

Introns
graphic display 211

Invitrogen
Custom Primers 72
custom primers 65
Customer support contact information 659
Gateway Cloning 321
Gateway Cloning from DB Explorer 186
ordering custom primers 72, 273, 284
proprietary molecules 165, 208
Ultimate ORF BLAST 290
Ultimate ORF Blast 290, 467
Website 660
Invitrogen proprietary 208

IUB Codes 667

Journal format 166

Keyboard shortcuts
Database Explorer 194

Keywords
- entering 171, 313, 356
- Molecule Editor 244
- object 171
- removing 171
- searching by in VNTI Database 174

L

Labels
- formatting 50

LabShare Database
- accessing in Vector NTI 639
- mini-exploring window in Vector NTI 639
- opening objects in VNTI 642
- Remote Database Sources Wizard 637
- saving objects from VNTI 642

Launching
- Analysis Monitor 560
- Analysis Monitor from GenomBench 542
- BLAST Search 575
- BLAST Search, Analysis Monitor 569
- GenomBench 506
- Query Application from GenomBench 545

License Manager 650, 687
- configuring dynamic license 654, 691
- configuring trial license 656, 693
- opening 650

License status 650, 687

License, static resetting 653, 691

Licenses, Vector Xpression 650, 687

Linear Display 209

Link Color
- in AlignX Blocks 599

Linker fragments 314

Linking panes 17
- tutorial 32

Linking/unlinking residues
- in AlignX Blocks 599

List Pane 132
- meaning of icons in 435
- Project Explorer 435

Loading
- Display Setup from a file 222, 225
- PCR analysis settings 247
- PCR settings 328

Local alignment
- in AlignX Blocks 593

Local Author 170

Low-complexity segments 579

LS Configuration utility 506

M

Managing
- shared database 386

Matrix
- BLOSUM 646
- PAM 646
- score 407
- Zoom in 412

Matrix Editor 646

Maximizing panes 16, 594

Memory cache 198

Methylation
- designating for RENs 187
- designating type for RENs 188
- showing in REN Database 187
- Text Pane Restriction/Methylation Map folder 188

Mini-Exploring window
- formatting page size 641
- navigating in 639
- opening LabShare objects 642
- saving objects to LabShare DB 642

Modifying
- Plot Setup 394
- VNTI DB Explorer window display 192
- VNTI window display options 204

Mol Fragments 316

Molar extinction coefficient calculation 665

Molecule
- adding features 236
- adding restriction sites to 236
- advanced Design, tutorial 93
- basic 305, 306
- changing starting coordinate 244
- changing starting coordinate, tutorial 42
- constructed 305, 306
- Construction 306
- Construction, tutorial 77
- create new using splicing 306, 307
- creating new 7, 306
- creating new from raw sequence 237
- defining complex Design fragments, tutorial 94
- defining construction fragments, tutorial 78
- defining Design fragments, tutorial 86
- Design 306
- Design, tutorial 85
- Display Setup 215
- documents 52
- editing a feature 240
- editing a sequence 240
- editing format 241
- editing graphics, tutorial 45
- Editing mode 48
- editing, Molecule Display window 233
- editing, tutorial 37
- effect of editing on parents 238
- exporting 232
- external files 206
- finding features in all panes 208
Formatting 52
Goal List 81
importing sequence text files 182
importing text file 182
making reverse complement 244
Molecule Editing mode 233
opening Molecule Shortcut file 206
parent-descendant relationships 669
Picture Editing mode 233
reconstruction 319
redesign 364
Revert To Saved command 40
shortcuts 232
viewing properties 195
Molecule Display window 8
description, tutorial 27
Display Setup 215
elements of 206
Graphics Pane 208
opening molecules 204
opening User Fields page 201
panes 206
resizing, tutorial 27
Sequence Pane 211
Text Pane 207
Text Pane, tutorial 31
viewing a highlighted region 210
working with sequence, tutorial 29
Molecule Document 232
description 205, 231
formats 8, 205
opening 206
Save As 52
saving 231
Molecule Editing mode 203, 233
for Construction/Design 308
Molecule Editor 305
Comments tab 243
description of 233, 237
editing a feature 240
for new molecule 238
General Information tab 238
imported text files in 182
Keywords tab 244
opening 237
Sequence and Maps tab 239, 242
User Fields tab 243
Molecule Feature dialog box 240
Molecule Shortcut file
creating 232
opening molecule from 206
Molecules 208
Invitrogen, proprietary 165
missing in database 322
Motifs
analyzing 217
editing 217
finding 217
in reports 377
removing 217, 224
saving 217
Sequence Pane display 211
setup for DNA/RNA 217
setup for protein 223
similarity 217, 218, 224
sorting 218, 225
Motifs search
launching 290
results 291
Motifs Setup 217, 223
Moving
fragment in a contig 485
object, Graphics Pane 228
Multiple alignment See Alignment, initiating multiple 402
Multiplex PCR
adding features 265
description 263
dialog box 266
initializing the analysis 266
Lists dialog box 264
process 264
report 270
Mutagenesis
launching search 291
restriction sites in 291
results 292
N
Name
searching by in VNTI Database 172
Naming
Design recombinants 361
of construction recombinants 318
NetWare
shared database 383
New Fragment dialog box 352
New Oligo
dialog box 285
non-ambiguous bases 137
Non-Cutter REN Analysis
dialog box 304
finding common non-cutters 302
finding common non-cutters, report 304
nucleotide sequences 154, 155
editing 146
nucleotide sequences, translating 144, 145
Object
applying new user fields 199
clearing user fields from 200
database, description 167
importing comments 170
viewing properties 195
Obtaining a Functional Entry Clone from Invitrogen 331
Oligo
Database Search 176
db operations on objects 185
searching by in VNTI Database 175
Oligo analysis
%GC method 665
dialog box 286
dimers 288
duplexes 289
hairpin loops 288
melting temperature 72
Nearest Neighbors method 287, 665
results 287
Tm 287
tutorial 71
Oligo Analysis dialog box 283
Oligo Database
searching for motifs 217
Oligo Duplexes
dialog box 283, 289
graphical depiction 288, 289
Oligo Editor 217, 284, 285
Oligo List
adding primer to 273
description 283
dialog box 283
loading oligos 217
opening 283
ordering custom primers 284
Online Help
description 13
Online Resources
Entrez Search System at NCBI 605
Open
DB Explorer 13
dialog box, modifying 195
Opening
3D-Mol 606
3D-Mol structures 606
AlignX 389
Analysis Monitor 560
BioAnnotator from VectorNTI 125
Construct dialog box 312
Display Setup 216
Dot Matrix 411
Fragment Window 465
Fragment Wizard 350
Gel Display window 365
GFF files 590
Goal Molecule Definition List 311, 355
hits as features, BLAST Viewer 590
Molecule Display window, tutorial 26
Molecule Document file 206
Molecule Editor 237
molecule from DB Explorer 205
molecule from shortcut file 206
molecule in display window 204
molecule with current profile 204
project in AlignX Blocks 598
project, BioAnnotator 422
Report Generation Wizard 376
Text Pane folders 208
Tools Manager 626
Vector NTI, tutorial 25
Opening at startup
DB Explorer 56, 164, 234
opening ContigExpress projects 132
Opening DB Explorer
startup 26
Options dialog box
3D-Mol, Graphics Pane 616
3D-Mol, Sequence Pane 620
ContigExpress 457
Vector NTI 204
VNTI Database Explorer 192
Order custom primers
Add to Oligo List 72
Ordering custom primers 72, 273, 284
ORFs 8
displaying in ContigExpress 144, 152, 153, 154
in reports 377
incomplete 221
nested 221
phase indicators 214
Sequence Pane display 212
showing in Contig window 478
translating 213, 293
translating, BioAnnotator tutorial 126
Ultimate ORF BLAST 290
undefined start/stop 221
ORFs folder 207
ORFs search
hiding results 290
launching 290
results 290
ORFs Setup 220
Overlapping 418

P
Page setup
header and footer codes 21
Pairwise alignmentSee Alignment-Pairwise 404
PAM 646, 662
Pane layout
modifying 22
Pane Manager 22
Panes
activating 15
linking 17
maximizing 16
re-sizing 16
switching between 15
Panes, maximizing 594
Parameters
Display Setup 215
multiple alignment 405
pairwise alignment 404
Parent/Descendant
Relationships 7, 669
Trees 185, 670
Parents
molecule, effect of editing on 238
status with copy feature 193
Password
shared database 385
Pasting
sequence 215
styles 230
PCR
analysis results 271
cloning sites box 272
create display window for product 272
finding product in Graphics Pane 272
product, creating display window 75
saving product in database 272
PCR Analysis
3' end tab 252
Alignment PCR 246, 260, 413
amplicon tab 249
Amplify Features dialog box 245, 247, 255
Amplify Selection dialog box 245, 247, 255
cluster tab, Multiplex PCR 269
dialog box, load/save buttons 247, 328
dialog boxes 245
displaying profile 274
features tab 257
filter tab 257
filters tab 254
Find PCR Primers dialog box 245, 247
Find Primers dialog box 66
Long PCR dialog box 245, 258
Multiplex PCR 246, 263
pairs tab 250
parameters 248, 255, 256, 259, 260
parameters, Alignment PCR 262
parameters, Multiplex PCR 267
primer melting temperature 67
primer tab 248, 255, 256, 259, 260
primer tab, Alignment PCR 262
primer tab, Gateway Cloning 325
primer tab, Multiplex PCR 267
printing PCR settings 274
printing results 274
qualities tab 253
report, Alignment PCR 263
report, Multiplex PCR 270
results tab, Alignment PCR 263
results tab, Multiplex PCR 268
results, tutorial 71
saving results to DB 273
similarity tab 251
structure tab 249
uniqueness tab 252
uniqueness tab, Multiplex PCR 269
Using Existing Oligos dialog box 245, 259
viewing saved results 274
PCR Analysis folder 207
PCR primers 246
adding restriction sites, tutorial 73
saving to database, tutorial 73
user defined 248
pDONR vectors
Entry Clones 323
Performing an LR Recombination 331
Pfam
parameters 565
updating Pfam database 559
Phase indicators
ORFs 214
translations 214
Phrap
importing Phrap file into ContigExpress 439
Phred
adding fragments to ContigExpress 439
loading quality values into ContigExpress 441
Picture Editing Mode 48, 203
activating 226
description 226
styles 229
Plot Pane
BioAnnotator 417
layout mode 418
modes in, BioAnnotator 417
overlapping plots 418
repositioning plots 418
resizing plots 418
resizing plots, BioAnnotator 419
Plot Setup
AlignX 394
Polylinker
editing for fragment trimming 499
poor consensus 151
Primer Similarity
dialog box 276
Primer/Oligo Parameters
dialog box 246, 277
Primer/Oligo Quality Specifics
dialog box 246
Primers
designing, reference 666
Order 273, 284
ordering custom 72, 273, 284
PCR, sense/antisense 272
similarity values 680
TaOpt 680
Tn values 680
Primers hybridization probes 279
Primers sequencing 275
Primers/probes
db operations on objects 185
Print 8
Print Preview
Dot Matrix 412
Printing
Display windows 21
from display windows, tutorial 33
from Gel Display window 374
from Project Explorer 459
PCR Analysis settings 274
reports 379
subset object list 184
Profile
open molecule using current 204
saving Display Setup 222
Profile alignmentSee Alignment, initiating profile 403
Project Explorer 432
Assembly Setup dialog 134
exiting 460
icons in lists 435
List Pane 132
List PaneSee List Pane 435
making reverse complement, CE 454
modifying viewing options, CE 457
opening a project 438
opening contigs/fragments 453
printing panes 459
saving a project 460
Status Bar settings 458
Status Bars 433
Tree PaneSee Tree Pane 433
tutorial 131
viewing formats 436
window 131
Project Explorer status 133
Project Explorer window 132, 133
Projects
Creating new 422
opening, BioAnnotator 422
Properties
opening 195
subset 181
Properties plots 662
Proprietary Molecules
Invitrogen 208
PROSITE
search and results, tutorial 111
ProSite
parameters 568
updating Prosite database 559
Protein
Analysis folder 208
analysis, tutorial 34
create display window, tutorial 35
create new by translation 306
create new using splicing 306, 307
creating new 186, 237, 306
Database Search 172
db operations on molecules 185
Display Setup 223
Feature Map Setup 223
horizontal resolution 225
molar extinction coefficient calculation 665
Motifs setup 223
Opening 34
parents 193
Sequence Setup 225
Protein analyses 662
Protein profile alignmentSee Alignment-Protein profile
406
Proteolytic Cleavage
parameters 566
PubMed/Entrez
Internet connection settings 633
search settings 632
PubMed/Entrez Search 631
Q
Quality specifics
primer/oligo 278
Query Application
configuring parameters 546
description 544
elements 545
Ensembl Genome Browser parameters 547
Ensembl SSAHA Search parameters 548
launching 545
PubMed Entrez Search parameters 548
UCSC BLAT Search parameters 546
UCSC Genome Browser parameters 546
viewing results 549
Query-Hit Alignment Graph 586
R

REBASE
- importing commercial source data 183
- importing enzymes from 182
- REBASE Enzymes dialog box 184
- web site 182

Recipient’s start position 313
Recombine an Expression Clone by BP
- Gateway Cloning 323
Redefine Style by Example 51
Redefining Styles 230
Redo 228

References
- in reports 377
- properties plots 663

Regular expression syntax
- anchors 684
- character sets 684
- description 683
- grouping 684
- metacharacter reference 685
- modifiers 684
- parts 683
- pattern memory 684

Remote Database Sources Wizard 636
- configuring Internet connection 636
- creating links to LabShare Database 637
- protocol options 638

Removing
- blocks in AlignX Blocks 596
- molecule from AlignX project 403
- motifs, all 217, 224
- motifs, selected 217, 224

Renaming
- database object 185
- subsets 181
- tool 626

RENs
- display settings 90, 96, 219
- Find Common Non-Cutting REN Analysis 302
- Non-Cutting REN Analysis 304
- Sequence Pane display 211

Replicon Type 313, 356

Report
- analyses 377
- authors 377
- comments 377
- description, General 375
- description, Relation 375
- Feature map 377
- formats 194
- general description 377
- generating 375
- Generation Wizard 375
- motifs 377
- options, selecting 376
- ORFs 377
- parent/descendant trees 380
- preparing, DB Explorer 193
- preview 379
- print preview 379
- printing 379
- references 377
- relation 380
- removing templates 381
- renaming templates 381
- restriction map 377
- saving as template 380
- selecting relationship information 381
- user fields in 377

Residue substitution matrix
- BLAST 580

Resizing
- display pane 16
- Molecule Display window 27
- plots in BioAnnotator 419

Resolution
- Prefer Constant Horizontal 225

Restriction enzyme
- cleavage points 187
- Database Search 176
- Find Common Non-Cutting REN Analysis 302
- methylase 187
- methylation points 187
- Non-Cutting REN Analysis 304

Restriction fragments
- calculating 300
- calculating, tutorial 105
- results 300
- selecting enzymes for 300

Restriction Map
- Rearrange Site Labels 33
- Restriction map 7
  - adding enzymes to 219
  - editing 242
  - in reports 377
- Restriction Map Setup 90, 96, 219
- Restriction Map/Methylation Map folder 207

Restriction sites
- adding to cloning fragments 314
- for Construction fragments, tutorial 81

Reverse complement
- making in AlignX 398
- making in ContigExpress 454
- making in VNTI 244

Revert to saved. 236

RFLP Analysis 300
- description 300
- dialog box 301
- results 301
results, filtering 302
results, performing gel analysis 302
showing fragment details 302
Run GW 264

S
Sample Fragment Properties 371
Save
Current Display Settings As 52
Molecule Documents 8
Save As
dialog box, modifying 195
Saving
3D-Mol results 624
3D-Mol, results 624
analyses, Analysis Monitor 571
BLAST Search results 589
Display Setup to a file 222, 225
items in project, CE 464
Molecule Document 232
molecule shortcuts 232
molecule to the database 232
PCR analysis settings 247
PCR primers, tutorial 73
PCR product, tutorial 75
project in ContigExpress 460
report templates 380
shared database connection 386
styles 230
subset object list 184
Schematic Pane
AlignX Blocks 597
Score Matrix
AlignX Blocks 602
Score matrixSee Matrix-Score 407
Scripts, configure 628
Scroll bars, using 16
Searching
find sequence, DNA 214
find sequence, protein 214
VNTI Database 58, 171
Secondary peaks
calling, chromatogram 501
initating calling in Project Explorer 456
Selecting
data in BioAnnotator 419
fragments in gel lane 372
Genetic Code 175
plots in BioAnnotator 418
sequence regions 213
sites and symbols 210
Selection Techniques
Broadcast Selection feature 19
Selection techniques 18
AlignX Blocks, alignment 598
CE Project Explorer 442
fragments for construction 235
Graphics Pane 234
modifying selection 235
modifying selection, tutorial 28
Molecule Display window, tutorial 28
tutorial 28
Separation time
Gel electrophoresis 373
Separator
description, Database Explorer 167
Sequence
changing text attributes 18
creating new molecule from 237
cut, copy, paste 215
display setup, BioAnnotator 421
Edit dialog box 240
editing 236
editing, tutorial 39
finding DNA 214
finding specific, CE 467
finding, protein 214
formatting text 18
insert new, tutorial 40
inserting 215, 236
selecting regions 213
viewing options, 3D-Mol 620
Sequence Pane
3D-Mol 611
BioAnnotator 420
copying data 20
custom numbering 212
feature display 211
Feature Map Setup, BioAnnotator 421
finding objects 17
Fragment Window 466
Molecule Display window 211
navigating 212
navigating, BioAnnotator 420
setting internal start point 212
translating in 293
viewing options, 3D-Mol 620
working in, tutorial 29
Sequence Profile graph 586
Sequence profile value 586
Sequence Setup
proteins 225
Sequencing Primers 275
dialog box 246, 275
folder 207
parameters 275
results 279
Similarity parameters 276
user defined 276
Sequencing primers
adding to oligo list 279
Set Caret Position dialog box 235
Set Selection dialog box 234
Setting
  genetic code 214, 294
  internal start point 212
  system attributes 186
Setup
  Dot Matrix 412
  Plot 394
Setup parameters
  AlignX Blocks 601
  contig assembly 444
Setup Profile
  saving 222
Shared database 383
  compatible file servers 383
  connecting to 384, 386
  creating new 383
  Database Explorer 386
  disconnecting from 387
  editing properties 386
  editing user accounts 386
  exchanging data with 387
  finding 385
  managing 386
  modifying objects 387
  networks 383
  passwords 385
  saving connection 386
Shortcut menus 14
Shortcuts
  in Database Explorer 194
  showing traces 469
Shrink Labels button 210
Shrink Symbols button 210
Silent mutations
  description 291
Sim4
  parameters 562
Similarity
  between ambiguous nucleotides 218, 680
  dialog box 246, 276
  primer parameters 276
Similarity Table
  BLAST Viewer 588
Sorting
  hits, BLAST Viewer 584
Source Content column 180
Source Subsets column 180
Spidey
  parameters 564
Spread Labels button 210
Standard Arrangement 33, 209, 231
  modifying 210
Starting
  Vector NTI, tutorial 25
  Starting coordinate
    changing 244
    changing, tutorial 42
Startup
  DB Explorer 26
  Opening DB Explorer 13, 204
  opening DB Explorer 56, 164, 234
Status Bar
  ContigExpress 462, 584
  display in DB Explorer 193
  settings in Project Explorer 458
Status bar
  Set Caret box 235
  Set Selection box 234
Storing
  graphics 20
  Stretch Labels button 210
  Stretch Symbols button 210
Structure Pane
  3D-Mol 609
  Alignments folder 621
  Edits folder 607
  information, 3D-Mol 609
Styles 52
  copy and paste 230
  defining 51
  loading 231
  modifying 230
  redefining 230
  saving 230
Subset
  adding objects to 179, 180
  clearing 181
  Create New 57
  creating new 179
  deleting contents 181
  description 167
  dismissing 181
  displaying properties 181
  editing 180
  exporting to text file directory 184
  importing molecules to 182
  printing object list 184
  removing molecules from 180
  renaming 181
  saving object list 184
  summary 181
Summary
  subset 181
Switch panes
  tutorial 65
Switching panes 15
System attributes, setting 186
System Information utility 656
Tables
  Citation 166
  Database Explorer, VNTI 166
TaOpt 680
Target sequences in GenomBench 521
Technical Support, contacting 659
Template
  AlignX Blocks 602
  saving report 380, 381
Template files
  in AlignX 409
Terminus Editor 317, 318
Text
  editing styles 227
  formatting sequence 18
Text file directories
  importing from 182
Text files
  importable formats 182
Text Pane
  AlignX Blocks 594
  Contig Window 474
  Design Description folder 362
  expanding/collapsing folders 16
  finding objects in other panes 17
  folders in 207
  folders, opening/closing 16, 208
  Fragment Window 466
  Gel Display window 370
  linking with other panes 17
  Molecule Display window 207
Thermodynamic Properties dialog box 283, 286
Tm
  %GC 287, 665
  %GC calculation 676
  ambiguous oligos 679
  general information 675
  Nearest Neighbors method 287, 665
  oligo analysis 287
  PCR primers 248
  primers/probes 680
  references 665, 681
  RNA oligos 679
  TaOpt 680
  Therm. Tm calculation 676
  thermodynamic 287
Toolbars 56
  switching panes, tutorial 65
Tools
  3D-Mol 620
  aligning chains, 3D-Mol 620
  calculating protein surfaces, 3D-Mol 622
  command field 627
  configuring 625
  creating new 626
  deleting 626
  description 627
  external, AlignX 413
  input formats 627
  name format 627
  Online Resources 605
  program-based vs script-based 628
  renaming 626
  script configure 628
  script field 627
  script-based 628
  with .tq extension 626
Tools Manager
  adding Web links 628
  opening 626
  sorting tools 628
TOPO cloning 321
  generating a clone 337
  high throughput cloning 347
  high throughput operations 344
  introduction 337
  PCR amplification 339
  traces, peak height 469
Translating
  CDS with splicing 294
  clearing Sequence Pane translations 213, 293
  codon usage table 294
  Complementary Strand 213
  Direct Strand 213
  DNA/RNA 213, 293, 306
  features 293, 294
  in six frames 293, 294
  in the Sequence Pane 213, 293
  into new protein molecule 213, 293, 306
  ORFs 213, 293
Translation
  Clear Translations 213, 293
  creating custom genetic code 295
  genetic code dialog box 295
  options 211, 213
  phase indicators 214
  setting genetic code 214, 294
  six frame 293, 294
  translate feature in Sequence Pane 211
  translate feature into new protein 211
  tutorial 30
Tree Pane
  Copy fragment/contig option 459
  History View mode 433
Trial License 650, 687
Trimming Fragments
  dialog box 494
  Ends Trimming dialog box 496
  vector contamination trimming 498
TrimmingSee Fragment trimming 498
internal start point 212
selected region 35, 210

Viewing options
- CE Project Explorer, modifying 457
- Contig Window 487
- Contig Window, Fragment Highlights tab 488
- Contig Window, General tab 487
- Contig Window, Weight Table tab 489

W
Weight Graph
- CE, calculation of 478
- CE, interpreting 477

Z
Zoom in
- matrix 411
Zoom to Selection 18
Zooming graphics
- BLAST Viewer 586
Zooming In/Out 17