Addressable configurations of DNA nanostructures for rewritable memory

Arun Richard Chandrasekaran*, Oksana Levchenko, Drhuv S. Patel, Molly MacIsaac and Ken Halvorsen*

The RNA Institute, University at Albany, State University of New York, Albany, NY, USA

ABSTRACT

DNA serves as nature’s information storage molecule, and has been the primary focus of engineered systems for biological computing and data storage. Here we combine recent efforts in DNA self-assembly and toehold-mediated strand displacement to develop a rewritable multi-bit DNA memory system. The system operates by encoding information in distinct and reversible conformations of a DNA nanoswitch and decoding by gel electrophoresis. We demonstrate a 5-bit system capable of writing, erasing, and rewriting binary representations of alphanumeric symbols, as well as compatibility with ‘OR’ and ‘AND’ logic operations. Our strategy is simple to implement, requiring only a single mixing step at room temperature for each operation and standard gel electrophoresis to read the data. We envision such systems could find use in covert product labeling and barcoding, as well as secure messaging and authentication when combined with previously developed encryption strategies. Ultimately, this type of memory has exciting potential in biomedical sciences as data storage can be coupled to sensing of biological molecules.

INTRODUCTION

The inherent nanoscale features and molecular recognition properties of DNA has made it useful for the construction of nanostructures with various applications in biology (1), medicine (2), materials science (3) and information processing and storage (4–6). As an archival storage medium, DNA is highly dense (~1 exabyte/mm³) (7) and long lasting (half-life of 500 years) (8), with recent efforts demonstrating storage of books and images (9) and a Shakespearean sonnet (10), and information retrieval of up to 215 petabytes per gram of DNA (11). However, archival storage systems are intended as ‘read-only’, necessitating the development of rewritable DNA systems providing short-term storage (12). The design of a memory device based on biomolecular interactions has been presented as early as the 1980s (13). DNA-based information processing systems reported so far include single and double stranded DNA (14) as bits ‘0’ and ‘1’, a hairpin-based memory stick with an address site on the loop (15), a three-state nanopatterned device providing eight possible memory states (16), and a translation system based on DNA double crossover (DX) tiles (17) (Supplementary Figure S1). Here, we present a user-friendly DNA-based memory system that can encode multiple bits of information with erasing, rewriting, write-protection and logic functionality.

Our memory system is based on encoding data in discrete conformational states of DNA nanostructures. To demonstrate the concept, we expanded upon previously developed DNA nanoswitches that exhibit binary switching behavior (18,19). The DNA nanoswitches self-assembled using DNA origami approaches (20) and purified from excess strands (21,22), have inducible loops that can be spatially programmed by placement of DNA overhangs at desired locations (23). In those previous works, the DNA nanoswitches were used as on/off sensors for detection and analysis of molecular interactions. In the simplest realization, DNA nanoswitches encode a single bit of information depending on the presence or absence of a single loop (1 or 0, respectively), detectable by gel electrophoresis (Figure 1). The loop is formed when single stranded overhangs on different sections of the nanoswitch are each partially hybridized with an external strand. For the purpose of this work, we will refer to the single stranded extensions as address sites and the external strands as data strands.

MATERIALS AND METHODS

Design and oligonucleotide mixtures

Oligonucleotides were purchased from Integrated DNA Technologies (IDT) with standard desalting. DNA nanoswitches were prepared as previously reported (18,19,23). In brief, the nanoswitch is a long duplex
Figure 1. DNA nanoswitch-based memory system. (A) Formation of the DNA nanoswitch. Lengths of the component DNA strands are shown in nucleotides (nt). (B) Concept used in this study: The DNA nanoswitch is a long duplex with single stranded extensions that are complementary to an external DNA strand. Hybridization of the data strand to these single stranded extensions results in the formation of a loop. (C) The linear (off) and looped (on) conformations of the nanoswitch can be read out using gel electrophoresis.

Comprised of a single stranded DNA scaffold (M13 viral genome) and short complementary backbone oligonucleotides. Among the backbone oligonucleotides, 12 strands (postal sites) are uniformly separated on the scaffold and can be replaced by address site oligonucleotides (illustrated in Supplementary Figures S2 and S3). Choosing different postal sites results in different loop sizes ('bits'). Address site oligonucleotides contain single stranded extensions that can bind to specific data strands.

A SnapGene DNA file with the backbone regions, postal and address sites labeled is provided as an additional Supporting file (in the SnapGene file, the bottom strand is the M13 sequence while the top strand represents the oligonucleotide sequences). The software SnapGene Viewer is available for free download at: http://www.snapgene.com/products/snapgeneviewer/

Linearization of M13 DNA

Five microliter of 100nM circular single-stranded M13 DNA (250 µg/ml, New England Biolabs), 2.5 µl of 10 × Cut Smart buffer (New England Biolabs), 0.5 µl of 100 µM BtsCI restriction-site complementary-oligonucleotide and 16 µl of deionized water were mixed and annealed from 95 °C to 50 °C in a T100™ Thermal Cycler (Bio-Rad, USA). One microliter of the BtsCI enzyme (20 000 units/ml, New England Biolabs) was added to the mixture and incubated at 50 °C for 15 min. The mixture was brought up to 95°C for 1 min to heat deactivate the enzyme followed by cooling down to 4°C.

Construction of nanoswitches

Linearized single-stranded M13 DNA (20 nM) was mixed with 10-fold excess of the backbone oligonucleotides, postal site oligonucleotides, address site oligonucleotides and filler strands. The mixture was annealed from 90°C to 20°C at 1°C min⁻¹ in a T100™ Thermal Cycler (Bio-Rad, USA). The constructs were either PEG precipitated (21) or LC-purified (22) after annealing to remove excess oligonucleotides. Purified constructs were diluted in 1 × PBS.

Nanoswitch operation and memory

The purified constructs (~400 pM) were mixed with desired concentration of the data strands (typically 2.5 nM final concentration) and incubated at room temperature (25°C). For the writing and erasing experiments, addition of data or eraser strands at subsequent steps was usually in 2× molar ratios. For the Hello World/Good Bye experiment (Figure 3F), data or eraser strands were added at 10-fold molar excess at each erase/rewrite steps. For time series experiments, specific data or eraser strands were added T–n hours preceding gel electrophoresis and at other shorter time intervals up until just before loading the gel (0 min).

Gel electrophoresis

Constructs were run in 0.8% agarose gels, cast from molecular biology grade agarose (Fisher BioReagents) dissolved in 0.5× Tris-borate EDTA (TBE) (Ultra pure grade, Amresco). Samples were mixed with a Ficoll-based loading solution. Gels were typically run at 75 V (constant voltage) at room temperature. Gels were pre-stained by mixing 1× GelRed stain (Biotium) with the gel solution before the gel was cast. Gels were imaged with a Bio-Rad Gel Doc XR+ gel imager and analyzed using the gel analysis tool in the Image Lab software package available with Bio-Rad Gel Doc XR+. For the Hello World → Good Bye multi-bit rewriting experiment, gels were run at 100 V (constant voltage) and imaged using a Typhoon 9400 variable mode imager (GE Healthcare). Image analysis was done using the software ImageJ (https://imagej.nih.gov/ij/). Median filter in Imagej was used to remove noise in gel images in Figures 2D, 3F and 4A–B.

RESULTS AND DISCUSSION

We extended the nanoswitch concept to build a multi-input memory system that can be triggered to produce different loops (Figure 2). This was achieved by programming the nanoswitch to contain address sites (analogous to single stranded overhangs used in (23)) on different locations along the scaffold (Supplementary Figures S2 and S3). By
reacting with partially complementary data strands, this arrangement was used to form loops of many different sizes and orientations. We characterized the effect of loop size and position on gel migration (Supplementary Figures S4 and S5 and Supplementary Tables S1 and S2) and chose five easily resolved loops to comprise a 5-bit memory system capable of encoding binary representations of alphanumeric characters using the Baudot code (24). This nanoswitch design has six address sites (spaced $\sim$600 nucleotides apart) that can form five unique loops using five different data strands (Figure 2A). The first address site is common among all five loops so that any given nanoswitch molecule can only form one loop representing one bit. As a result, this comprises a 5-bit memory system where each bit is independently addressable and resolved on an agarose gel (Figure 2B and C). We comprehensively tested the response of the system by demonstrating all 32 possible outcomes (Figure 2D and Supplementary Figure S6).

To add erasing and rewriting functionality to the system, we incorporated data strands with short single-stranded extensions to facilitate data strand removal by toehold-mediated strand displacement (25) (Figure 3A). Using this scheme, we demonstrated several writing and erasing cycles for the same bit or for different bits (Figure 3B and C and Supplementary Figure S7). We characterized concentration dependent writing and erasing speeds, which occur on the scale of minutes to hours for sub-$\mu$M to sub-nM concentrations of external strands, respectively (Supplementary Figures S8–S10). We also found that different bits have different write speeds, which are likely to be dependent on loop size (smaller loops generally form faster due to proximity of address sites) and binding sequence. Under ideal conditions, erasing is complete with no residual band from the first writing event, and the second writing does not suffer significant signal degradation. Erasing of pre-written bits can be tuned by the ratio of eraser to data strands used and recycling of information over several write/erase cycles is also feasible using this system (Supplementary Figure S11). Multiple write/erase cycles will require ever increasing concentrations, posing practical limitations on the maximum cycle number as strand competition, kinetics, and upper concentration limits come into play. Based on our current rewriting approach and our previously published detection range spanning $\sim$5 orders of magnitude in concentration (23), the practical upper limit is probably in the 10–100 cycle range (<1.8 fold molar excess at each step, 100 cycles corresponds to $<1.06$ fold molar excess at each step). The writing and erasing speeds of this system could potentially be improved by accelerating hybridization rates (26) between address sites and data strands and

Figure 2. Design and operation of the memory system. (A) Design of the 5-bit memory system with multiple address sites (A0–A5) that are partly complementary to the data strands (D1–D5). Binding of different data strands on the address sites yields loops of varying sizes. (B) A scheme showing the operation of the memory system yielding different outputs (bits) for different data strands. For example, the grey and blue halves of data strand D1 bind to address sites A0 (grey) and A1 (blue) respectively, thereby forming a loop. (C) Gel read-out of the 5-bit memory system. The first lane contains only the nanoswitch with no data strands added (no loop formation). (D) Gel readout of all possible bit combinations using the 5-bit memory system (image cropped and adjusted for clarity; untouched image and explanation in Supplementary Figure S6).
Figure 3. Erasing and rewriting capability of the memory system. (A) Illustration of the toe-hold strand displacement-based erasing and rewriting. (B) An example showing writing, erasing and rewriting of bit-3. (C) An example showing writing and erasing of bit-3 followed by writing of a different bit (bit-2) on the same system. (D) A write-protection strategy using data strands without toe-holds. Specific bits can be write-protected while others can be erased. (E) Gels showing controlled write-protection of bit-3 or bit-2. (F) Demonstration of multi-bit write-erase-rewrite.

by choice of component DNA sequences (27). It is worth noting that erasing and rewriting do not occur without the presence of the toehold on the data strands (Supplementary Figure S12). This feature acts as a selectable ‘write protection’ function, enabling either ‘read only’ or ‘read and write’ functionality (Figure 3D and E). We showed multi-bit information processing by encoding the words ‘Hello World’ using the Baudot code, erasing the information and rewriting ‘Good Bye’ on the same system (Figure 3F).

Next, we demonstrated logic-gated response of our system by designing a single loop that can be triggered by two input strands (Figure 4). Some examples of previous nucleic acid-based logic systems include logical circuits that operate in cells (28), strand displacement-based chemical controllers (29), and a deoxyribozyme-based molecular calculator (30). To demonstrate an OR gate, we programmed two address sites each of which have regions complementary to part of input strand A (orange) and a region complementary to part of input strand B (light blue). Addition of either input causes loop formation by hybridization of input strand A or B onto the address sites (Figure 4A). For the AND gate, we designed the system to contain only one address site (orange) that was complementary to part of input strand A. One half of input strand B (light blue) was designed to bind directly on the scaffold strand and the other half to input strand A. The loop is triggered only in the presence of both inputs, being held together by the hybridization of the complementary regions on A and B (Figure 4B). The use of our method can be expanded to include other logic gates by designing specific binding partners for the input strands and the address sites. A given state of nanoswitch can be stabilized using triplex forming oligonucleotides as latches (31), that bind to the duplex formed between the address sites and the input strands (by choosing triplex specific sequences for the address sites and input strands). Moreover, incorporating toehold-containing input strands and/or address sites could provide a strategy to design reversible (32) or re-settable (33) logic gates.

While we showed a 5-bit system here to encode binary representations of letters, our overall scheme offers some scalability. We successfully tested an 8-bit system (Supplementary Figure S13), and with some optimization the system could potentially be scaled to dozens of bits (several bytes). While this level of data storage is miniscule compared with archival DNA systems, it could prove useful for certain applications such as product labeling. The stored information is recoverable after drying (Supplementary Figure S14), suggesting that products could be discreetly
Figure 4. Logic-gated response of the memory system. (A) Operation, truth table and results of an OR gate. The nanoswitch contains two address sites each of which have regions complementary to part of input strand A (orange) and part of input strand B (light blue). Addition of either input causes loop formation. (B) Operation, truth table and results of an AND gate. The nanoswitch contains one address site (orange) that is complementary to part of input strand A. One half of input strand B (light blue) binds to a single stranded region on the scaffold (indicated by arrow) and the other half to input strand A. The loop is triggered only in the presence of both inputs.

We have shown proof of concept for a simple and effective method of rewritable memory storage using DNA nanostuctures and data readout using gel electrophoresis. The gel readout is rather simple and inexpensive when compared to other readout methods that have been used including FRET (16), optical outputs (35), electrochemical readout (36) or atomic force microscopy (AFM) (37). In previous work, we demonstrated resolution of a looped nanoswitch in as little as 10 minutes (23), and such a readout could be implemented outside of a lab using bufferless gel systems currently available (e.g. ThermoFisher E-gel). Still, our system could be compatible with different types of readouts including single-molecule methods such as nanopore (38), AFM (37), and centrifuge force microscope (CFM) (39). One advantage of single-molecule readouts is the possibility to read multiple bits per molecule rather than the one bit per molecule that we used. Multiple loops in one nanoswitch molecule are possible (19), but tend to cause complex gel banding patterns that make readout difficult. These problems could be overcome if direct single-molecule probing was employed, and could dramatically increase the storage capacity. This could additionally facilitate expansion of bit depth that would be needed, for example, in Universal Product Code (40 bits).

Our memory system is also compatible with an encryption scheme that we previously presented (14), giving it further applications in secure messaging, authentication, and anti-counterfeiting. In that encryption scheme, the data strands act as a decryption key for the prepared nanoswitches. Without access to the physical mixture that comprises the decryption key, the data cannot be retrieved. Since the sequences of the data strands can be arbitrarily designed and kept secret, encrypted data remains secure even when distributed publicly. This data encryption scheme is highly asymmetric; simple for authorized users to encrypt and decrypt data but extraordinarily difficult for unauthorized decryption. Simple countermeasures can further strengthen the encryption (40), which unlike conventional techniques is not directly threatened by increasing computational power.

Some of the most exciting potential applications are in biotechnology, where the sci-fi notion of molecular robotics complete with sensing and actuation are quickly becoming a reality (41,42). Our memory system operates based...
on molecular recognition of nucleic acid sequences, and the concept could be adapted for other biological molecules by integrating aptamers or conjugated antibodies. Memory manipulations and simple decision-making could be performed with biological inputs or outputs. This type of memory could potentially be integrated with molecular robotics, especially DNA origami structures, to integrate data storage or computation with sensing or actuation.

SUPPLEMENTARY DATA
Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS
We thank The RNA Institute (SUNY Albany) for financial support. We also thank Prof. Wesley Wong (Wyss Institute at Harvard) and Dr Mounir Koussa (Confer Health, Inc.) for comments on the manuscript. A.R.C. thanks Confer Health, Inc. for providing lab space to perform the multibit rewriting experiments.

Author contributions: A.R.C. and K.H. conceived the project, designed experiments, analyzed data and wrote the manuscript. A.R.C., O.L., D.S.P. and M. M. performed experiments.

FUNDING
NIH/NIGMS [R35GM124720 to K.H.]. Funding for open access charge: NIH/NIGMS [R35GM124720 to K.H.].

Conflict of interest statement.

REFERENCES