Toward Catalytic Networks from Thermodynamic Binding Networks

Keenan Breik[1], Cameron Chalk[1], David Doty[2], David Haley[2], and David Soloveichik[1]

[1] University of Texas at Austin
[2] University of California, Davis

The formal Thermodynamic Binding Network (TBN) model [2,3] shows that complex chemical behavior can arise solely from the thermodynamic driving forces of bond formation and the configurational entropy of forming separate complexes. Are these forces sufficient to induce arbitrary desired kinetic behavior by programming the energy landscape?

A catalyst enables a chemical reaction without getting consumed in it. Catalysis is one of the more complex chemical behaviors, and it occurs widely in biology. There are large catalytic networks (e.g., transcriptional, phosphorylation) that are dedicated to chemical information processing, and that regulate all aspects of the cell's behavior.

Our previous work developed a TBN implementing a single catalytic reaction $C + H \rightleftharpoons C + V$ [1]. A *template* molecule can take on two states H and V depending on which other molecules bind to it. Without an additional *catalyst* molecule, there is an energy barrier to change a template in state H to state V. This uncatalyzed energy barrier is entirely due to many separate complexes having to come together to achieve the transition, and can be made arbitrarily large by increasing the size of the monomers. Conversely, when the catalyst is present, there is a path with the number of separate complexes never decreasing by more than one.

Toward the goal of implementing networks of catalytic reactions, we extend the prior method in two distinct ways. Each method allows designing a template molecule to take on arbitrarily many states, with specific catalysts between any two of the states. The first method uses a number of binding sites that is exponential in the number of states and retains the full uncatalyzed energy barrier. The second construction uses concepts from the famous social golfer problem in combinatorics to reduce the number of binding sites to a quadratic while retaining an uncatalyzed energy barrier proportional to that of the first construction.

Our work leaves a number of open questions. First, it is open whether the energy barrier of the quadratic construction remains arbitrarily programmable if multiple catalysts are present simultaneously (this is proven true for the exponential construction). Second, it is unclear whether either construction can be extended to the case of multiple template molecules being simulataneously present. Finally, we ask whether our techniques can be generalized to simulate arbitrary catalytic networks in which a species can be both a template and a catalyst in different reactions. Given the generality of the TBN model, our work could suggest design strategies for preventing undesired kinetic behavior in a variety of molecular systems, including autocatalysts (which are a special case of catalytic networks), and nucleation barriers.

References

1. Keenan Breik, Cameron Chalk, David Doty, David Haley, and David Soloveichik. Programming substrate-independent kinetic barriers with thermodynamic binding networks. In *International Conference on Computational Methods in Systems Biology*, pages 203–219. Springer, 2018.
2. Keenan Breik, Chris Thachuk, Marijn Heule, and David Soloveichik. Computing properties of stable configurations of thermodynamic binding networks. *Theoretical Computer Science*, 2018.
3. David Doty, Trent A. Rogers, David Soloveichik, Chris Thachuk, and Damien Woods. Thermodynamic binding networks. In Robert Brijder and Lulu Qian, editors, *DNA Computing and Molecular Programming: 23rd International Conference*, pages 249–266. Springer, 2017.

DNA-Encircled Lipid Bilayers

Katarina Iric,[a,b,c] Madhumalar Subramanian,[b,c] Jana Oertel,c Nayan P. Agarwal,[a] Michael Matthies,[a] Xavier Periole,[d] Thomas P. Sakmar,[e] Thomas Huber,[e] Karim Fahmy*[b,c] and Thorsten L. Schmidt*[a,f]

[a] *Center for Advancing Electronics Dresden (cfaed), Technische Universität Dresden 01062 Dresden, Germany.*
[b] *Helmholtz-Zentrum Dresden-Rossendorf, Institute of Resource Ecology, 01328 Dresden, Germany*
[c] *Biotechnology Center, Technische Universität Dresden, 01307 Dresden, Germany*
[d] *Department of Chemistry, Aarhus University, Aarhus C, Denmark*
[e] *Laboratory of Chemical Biology & Signal Transduction, Rockefeller University, NY 10065 New York, USA*
[f] *Department of Physics, Kent State University, Kent, OH 44242, USA*

Email: tschmi21@kent.edu

Lipid bilayers and lipid-associated proteins play crucial roles in biology. As in vivo studies and manipulation are inherently difficult, membrane-mimetic systems are useful for the investigation of lipidic phases, lipid-protein interactions, membrane protein function and membrane structure in vitro. In this work, we describe a route to leverage the programmability of DNA nanotechnology and create DNA-encircled bilayers (DEBs). DEBs are made of multiple copies of an alkylated oligonucleotide hybridized to a single-stranded minicircle, in which up to two alkyl chains per helical turn point to the inside of the toroidal DNA ring. When phospholipids are added, a bilayer is observed to self-assemble within the ring such that the alkyl chains of the oligonucleotides stabilize the hydrophobic rim of the bilayer to prevent formation of vesicles and support thermotropic lipid phase transitions. The DEBs are completely free of protein and can be synthesized from commercially available components using routine equipment. The diameter of DEBs can be varied in a predictable manner. The well-established toolbox from structural DNA nanotechnology, will ultimately enable the rational design of DEBs so that their size, shape or functionalization can be adapted to the specific needs of biophysical investigations of lipidic phases and the properties of membrane proteins embedded into DEB nanoparticle bilayers.

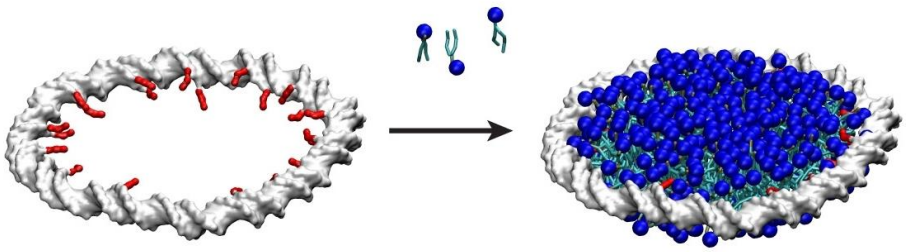


Figure 1: Reconstitution of a lipid bilayer inside an alkylated DNA minicircle.

Reference

[1] Iric, K.; Subramanian, M.; Oertel, J.; Agarwal, N. P.; Matthies, M.; Periole, X.; Sakmar, T. P.; Huber, T.; Fahmy, K.; Schmidt, T. L. DNA-Encircled Lipid Bilayers. *Nanoscale* **2018**, *10* (39), 18463–18467.

A Programmable DNA Nano-indicator-based Platform for Large-scale Square Root Logic Biocomputing

Chunyang Zhou,[ab] Pengfei Wang,[ab] Yonggang Ke[a]

[a] Institute of Molecular Medicine, Renji Hospital, Shanghai Jiao Tong University School of medicine, and School of Chemistry and Chemical Engineering, Shanghai Jiao Tong University, Shanghai, 200120, China.

Email: pengfei.wang@emory.edu

[b] Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, Georgia, 30322, United States

ABSTRACT: DNA computing is one of the most promising candidates for next-generation molecular computers. However, most of DNA logic circuits are still incapable of realizing complex mathematical operations, such as the construction of square root logic operations, which can be carried out with a four-bit binary number. In this work, by utilizing the powerful DNA hybridization and toehold-mediated DNA strand displacement, a large-scale 10-bit square root logic operation was constructed, which can calculate the square root of a 10-bit binary number (within the decimal number 900).

Superior to previous logic operations, this large-scale 10-bit square root logic circuit could calculate the binary number within 10 bit and output the five distinct fluorescence signals to make sure the calculation results. This logic operation is based on a series of robust and smart nano-indicators, which can serve as the reacting platform to construct the large-scale 10-bit square root logic operation. This nano-indicators perform a perfect low background, which is crucial for coding and implementing DNA algorithms.

This work not only dramatically opened the potential of DNA-based biocomputing with large-scale input data, but also unveiled novel horizons for biomaterial-based logic devices, logic-controlled biosensing, bioimaging, genome editing and so on.

DNA nanostructure-based label-free detection of cancer miRNA-21 biomarker

Shuo Han[1], Shuo Yang[1], Wenyan Liu[2]* Risheng Wang[1]*
[1] Department of Chemistry, Missouri S&T, Rolla, MO 65409, USA

[2] Center of Research for Energy and Environment, Missouri S&T, Rolla, MO 65409, USA

Contact: wangri@mst.edu, liuweny@mst.edu

MicroRNAs (miRNAs) have emerged as the promising molecular biomarkers for early diagnosis, and enhanced understanding of the molecular pathogenesis of cancers as well as certain diseases. Here, a facile, label-free, and amplification-free electrochemical biosensor was developed to detect miRNA by using DNA origami nanostructure-supported DNA probes, with methylene blue (MB) serving as the hybridization redox indicator.

In this design, DNA origami was utilized, for the first time, as the miRNA sensing platform to provide a large number of free-standing DNA probes for capturing target nucleic acids with enhanced probe accessibility. In addition, the DNA origami probes were immobilized on the working electrode surface via electrostatic adsorption between a cationic and biodegradable chitosan film and negatively charged DNA origami nanostructures. This method offered the advantage of, not only providing a simple, yet stable attachment of DNA origami on the electrode surface, but also avoiding the use of modified thiol-DNA or biotin-DNA. With a facile fabrication method that does not require laborious labeling, probe immobilization, and signal amplification, our strategy promises the aforementioned advantages in determining the concentration of miRNA.

The successful immobilization of DNA probes (ss-DNA) and its hybridization with targeted miRNA-21 molecules were confirmed by electrochemical impedance spectroscopy (EIS) and cyclic voltammetry (CV) methods. This strategy employed for determination of target miRNA-21 in cancer exhibited high sensitivity. Its linear detection range was from 0.1pm to 10nM with a lower detection limit of 78.3 fM. The selectivity of the miRNA biosensor was also studied by observing the discrimination ability of a non-complementary target. Overall, this strategy has demonstrated great potential for sensitive, selective and label-free determination of miRNA in cancer diagnosis at the early stage, and prognosis of certain diseases.

Dynamic DNA materials powered by artificial metabolism

Shogo Hamada and Dan Luo
Department of Biological and Environmental Engineering, Cornell University
hamada@cornell.edu, dan.luo@cornell.edu

Metabolism is one of the important characteristics of life. Life dynamically and hierarchically generates materials by a flux of matter and energy. By mimicking this mechanism, here we developed dynamic DNA materials powered by artificial metabolism[1]. Biochemical synthesis/decomposition and dissipative assembly of DNA were coupled as one material system. Anabolism allowed dynamic and hierarchical generation of mesoscale patterns from nanoscale building blocks. By integrating catabolism, the patterns were autonomously generated, degenerated, and then also regenerated. Various applications of the material, from DNA-based machines to pathogen detection, were demonstrated. Notably, machines with emergent locomotion behavior were programmed using FSA (Finite State Automata). A continuous generation at the head of the body and the degeneration at the tail implemented the behavior. Furthermore, by expanding the program, a racing behavior of two competing bodies was also achieved. The machines could be used as a locomotive element of molecular robots. Artificial metabolism could open a new route to creating dynamic biomaterials with life-like characteristics.

[1] Shogo Hamada*, Kenneth Gene Yancey, Yehudah Pardo, Mingzhe Gan, Max Vanatta, Duo An, Yue Hu, Thomas L. Derrien, Roanna Ruiz, Peifeng Liu, Jenny Sabin, Dan Luo*. Dynamic DNA material with emergent locomotion behavior powered by artificial metabolism. *Science Robotics*, 4(29), eaaw3512 (2019).

A DNA-Encoded Spatial Sender-Receiver System

Shuo Yang, Alex Joesaar, Tom F.A. de Greef
Eindhoven University of Technology

The molecules of a cell, and at a higher length scale, cells in an organism are nodes in an interwoven communication network, collectively transmitting and receiving, coding and decoding information. Using a constructive approach, synthetic biologists have started to unravel the design principles behind intercellular communication using genetically modified cells. However current experimental tools are insufficient to understand the behaviour comprehensively and quantitatively. Therefore, the design and construction of simple model systems featuring compartmentalised signalling is highly desirable.

Previously our group has developed an artificial communicating system through a bottom-up approach, where protein-polymer microcapsules act as synthetic protocells and short DNA oligomers function as signalling molecules.[1] On the basis of this system, we are keen to explore the potential of spatial controlling protocellular communication. In this present work, we report a DNA-encoded spatial sender-receiver system. Specifically, in a microfluidic trap array, the signal molecules diffuse out from a single sender after receiving a light stimulus and activate the surrounding receivers through DNA strand exchange. We investigate receivers with different characteristics, such as the internal concentration of DNA template, the permeability of the protocellular membrane and catalytic regeneration of signal DNA. Then environmental effects such as the density of receivers and extracellular degradation of DNA signals are investigated. In addition to these, a system with spatial logic AND gate is developed. Our work shows that DNA-encoded sender-receiver systems present a simple and versatile platform for mimicking sender-receiver architectures in living systems.

1. Alex, J. et al. DNA-based communication in populations of synthetic protocells. *Nat. Nanotechnol*. **14**, 369-378 (2019).

Engineering Kinetically Uniform Sets of DNA Sequences

Michael Tobiason[1], Bernard Yurke[1,2], and William L. Hughes[1,*]

[1] Micron School of Materials Science & Engineering, Boise State University, Boise, Idaho, 83725, USA
[2] Department of Electrical & Computer Engineering, Boise State University, Boise, Idaho, 83725, USA

[*] To whom correspondence should be addressed.
Tel: 208.426.4859; Email: WillHughes@BoiseState.edu

ABSTRACT

DNA duplex-formation and strand-displacement reactions are fundamental to the field of DNA nanotechnology. However, the relationship between DNA sequence and the rate of these reactions is not fully understood. It is hypothesized here that observed kinetic variations arise due to unintentional base-pairing and that by eliminating such interferences kinetically uniform sets of sequences can be engineered. By analyzing 242 unique sequence/rate pairs from the literature, it is shown that ranking sequences based on the size and number of their largest unintentional intramolecular complements reliably identifies statistically-significant (p-values < 0.05) sub-populations with improved kinetic reproducibility. To engineer kinetically uniform sequences, the Sequence Evolver software tool and underlying Total-Fitness-Score are presented. By characterizing the reaction rates of twelve sequence-sets at a range of temperatures, it is experimentally demonstrated that kinetic reproducibility increases as the length of the longest intramolecular interference decreases. The highest level of reproducibility is observed when all intramolecular interferences 3 base-pairs or longer are eliminated. For such sequence-sets, the observed rate variations of ± 9% and ± 2% establish new standards in kinetic reproducibility for duplex-formation and strand-displacement reactions, respectively. Based on further analysis of the 242 unique sequence/rate pairs, it is speculated that variation in the hybridization rates of such sequence-sets will remain less than ± 10% across a range of experimental conditions. The performance of the SeqEvo software as a sequence-generation method is benchmarked against five alternative tools, verifying that the quality of SeqEvo sequences represents an improvement relative to current state-of-the-art methods. Furthermore, the versatility of the SeqEvo software is demonstrated by utilizing it to generate new sequences for four previously-published systems with novel architectures including: the 10x10x10 DNA brick published by Ke *et al*. (1), a seesaw-gate based network published by Qian *et al.* (2), the autocatalytic four-arm junction published by Kotani et al. (3), and the autocatalytic network published by Zhang *et al.* (4).

References

1. Ke, Y.G., Ong, L.L., Shih, W.M. and Yin, P. (2012) Three-dimensional structures self-assembled from DNA bricks. *Science*, **338**, 1177-1183.
2. Qian, L. and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, **332**, 1196-1201.
3. Kotani, S. and Hughes, W.L. (2017) Multi-Arm Junctions for Dynamic DNA Nanotechnology. *J Am Chem Soc*, **139**, 6363-6368.
4. Zhang, D.Y., Turberfield, A.J., Yurke, B. and Winfree, E. (2007) Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, **318**, 1121-1125.

Developing Privileged Genetic Alphabets for DNA Information Storage

Raghav Shroff, Jared Ellefson, Alex Boulgakov, Randy Hughes, Edward Marcotte, and Andrew D. Ellington
Center for Systems and Synthetic Biology
University of Texas at Austin
Austin, TX

The possibility of deep or archival data storage with DNA is increasingly becoming a reality, in part due to improvements in the read-write infrastructure, both at the molecular and at the systems level. While DNA as a storage medium has many advantages, it is also possible that storage capacity could be improved or extended by using XNAs. For example, many XNAs are far less susceptible to nuclease degradation, which would improve the prospects for longer term information storage, and reduce requirements for redundancy. In order to use XNAs at scale, it would be useful to have enzymes that could readily undertake their reading and writing, and that could be simply adapted to larger systems and algorithms for DNA data storage that may yet be developed. To this end, we have evolved a single enzyme that can readily read 2'-OMe DNA, and then write DNA for sequencing; this enzyme should allow extremely stable 2'-OMe oligonucleotides, which are highly resistant to nuclease degradation, to serve as the medium for longer term storage. Beyond reporting on the generation and characterization of this extremely novel enzyme, we have encoded a variety of files (such as the Wikileaks Mission Statement, some of the original Enigma encoded messages, and a tortilla recipe) into 2'-OMe DNA and showed that we can recover the encoded information using a single enzyme, the 'privileged' polymerase. In addition to the 2'-OMe-utilizing enzyme, we have generated a variety of other enzymes that can read and write XNAs, including RNA polymerase variants that can be used to generate 2'-OMe DNA, thus completing at least one enzymatic set for both storage and recovery of very stable information. We have also developed RNA polymerase variants that can incorporate novel base-pairs that expand that genetic alphabet (i.e., Hachimoji DNA), and these may also prove of utility in expanding the capabilities for (and information density of) XNA data storage.

Title: DNA-based logic gates with alphanumeric readout for the analysis of clinically relevant point mutations

Ryan P. Connelly and Yulia V. Gerasimova

Chemistry Department, University of Central Florida

Molecular logic gates analyze chemical inputs to produce binary outputs, mimicking the logical operations of traditional electronic logic gates. DNA logic gates have been suggested for the design of a DNA-based microprocessor; due to biocompatibility, such structures may be advantageous for monitoring and attenuation of processes within biological systems. While DNA-based microprocessors are a long way from realization, DNA-based logic gate technology can be already applied in diagnostics. For example, by analyzing nucleic acid biomarkers (e.g. miRNA) or genetic signatures of bacterial and/or viral pathogens, logic gate based sensors can help diagnose cancer or infectious diseases. In case of DNA-based logic constructs with a capacity to differentiate sequences down to single-nucleotide substitutions, antibiotic resistance profiles of pathogenic bacterial can be determined to ensure proper treatment of bacterial diseases. By incorporating a logical element directly into a sensor system, computation and analysis can be performed without the need for expensive instrumentation or smart devices.

In this work, we explored a multicomponent (or split) approach for the design of DNA-based logic gates with visual output. High output results from either a catalytic function of a G-quadruplex (G4) peroxidase-like deoxyribozyme (Dz) (as in Fig. 1), or the assembly of a dye-binding DNA aptamer. We designed the constructs to exhibit either positive or negative logic. Positive logic is achieved by splitting the input-recognition module of the logic gate into two halves, each of which is elongated with a half of the signal-transducer module (Fig. 1A). When a correct input is present, the signal transducer is formed, and high output is observed. In case of negative logic (Fig. 1B), the signal transducer module is pre-formed in the absence of the input. In the presence of a correct input, this structure disassembles to yield low output.

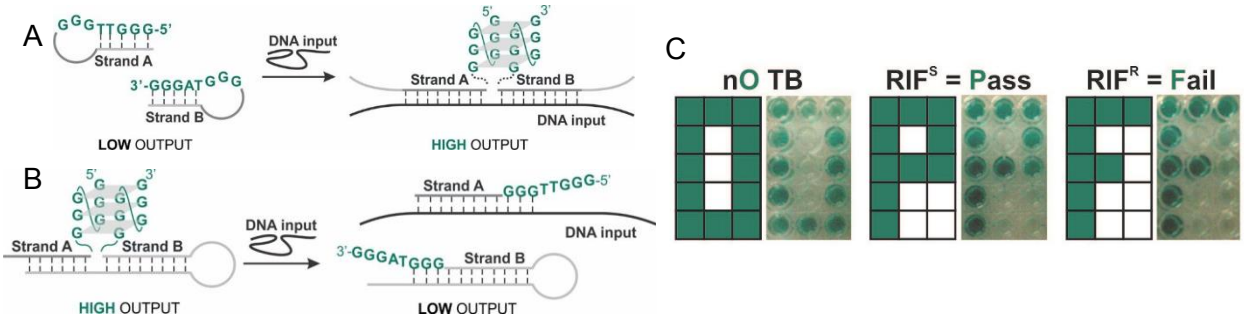


Figure 1. Multicomponent DNA logic gates based on G4 peroxidase-like Dz exhibiting positive (A) or negative (B) logic. C. Visual display for rifampin (RIF) susceptibility testing. Display images for no pathogen ("no TB"), or pathogen of RIF-sensitive ("RIFS") or resistant ('RIFR") genotype.

We arranged a set of the designed logic gates into a well-plate to create visual alphanumerical displays (Fig. 1C) for drug-resistance profiling of *Mycobacterium tuberculosis*. Alphanumeric readouts may reduce the processing of several logic gates into a single character that can be read without instrumental assistance.

Funding from the National Institute of Health (NIH; 1R21AI123876-01A1) is greatly appreciated.

Engineering RNA-Sensing CRISPR-Cas Systems

Mike Jin[1,2], Nicolas Garreau de Loubresse[1,2], Youngeun Kim[1.2], Jongmin Kim[1.2.3], and Peng Yin[1,2]

[1] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, United States

[2] Department of Systems Biology, Harvard University, Boston, Massachusetts 02115, United States

[3] Division of Integrative Biosciences and Biotechnology, Pohang University of Science and Technology, Pohang, Gyeongbuk 37673, Republic of Korea

The programmability of CRISPR-derived Cas9 as a sequence-specific DNA-targeting protein has made it a powerful tool for genomic manipulation in biological research and translational applications: only a single-guide RNA (sgRNA) is required for its activity, and its sequence specificity is easily programmable with sgRNA complementarity to the desired DNA target. To fine-tune the on-target activity of Cas9 as well as to mitigate off-target effects, many strategies have been developed to regulate Cas9 activity by post-translational regulation of Cas9, including using exogenous activating triggers such as light or small molecules. Cas9 activity can be more programmably engineered to respond to nucleic acids, but these efforts have focused primarily on single-input control of Cas9, and until recently, they were limited by sequence dependence between parts of the guide RNA and the sequence to be detected. Here, we not only design and present DNA- and RNA-sensing conditional guide RNA (cgRNA) that have no such sequence constraints, but also demonstrate a complete set of logical computations using these designs on DNA and RNA sequence inputs, including AND, OR, NAND, and NOR. The number of RNA logical inputs can also be scaled up using the self-assembling properties of RNA triggers so that a single switchable sgRNA can sense multiple RNA inputs. If applied in cellular contexts, sensing endogenous signals such as transcriptomic RNA signatures could enable improved spatiotemporal control by limiting CRISPR activity to biological contexts (e.g. tissues, cell types) with the desired transcriptomic profile, thus minimizing undesired off-target effects outside of these contexts. The development of sequence-independent nucleic acid-sensing CRISPR-Cas9 systems with multi-input logic computation capabilities could lead to improved genome engineering and regulation as well as the construction of synthetic circuits with broader functionality.

A pH-Responsive DNA Walker that is Dynamically Controlled

Dongbao Yao[1], Sanchita Bhadra[2], Erhu Xiong[2], Haojun Liang[1], Andrew D. Ellington[2] and Cheulhee Jung[3*]*

[1]CAS Key Laboratory of Soft Matter Chemistry, iChEM (Collaborative Innovation Center of Chemistry for Energy Materials), Department of Polymer Science and Engineering, University of Science and Technology of China, Hefei, Anhui 230026, P. R. China
[2]Institute for Cellular and Molecular Biology, Department of Molecular Biosciences, University of Texas at Austin, Austin, Texas 78712, United States
[3]Division of Biotechnology, College of Life Sciences and Biotechnology, Korea University, Seoul 02841, Republic of Korea

E-mail: ellingtonlab@gmail.com; damo363@korea.ac.kr

For most DNA walkers that have been reported so far, the walking rate is fixed once walking is initiated. This is because association and dissociation of walkers from tracks are for the most part non-programmable (with the notable exception of a light-controlled DNA walker). In response to these limitations, we designed a system whose movement responded to alternating pH changes stimulation and that could potentially be controlled by pH gradients. For this purpose, we chose a well-characterized, pH-responsive CG-C+ triplex DNA[1] that forms through Hoogsteen interactions between cytosine in the triplex-forming strand (TFS; pKa of cytosines in triplex DNA are ~6.5)[2] and G:C pairs in the duplex. Triplex formation was then adapted to our previously designed, tetrameric catalytic hairpin assembly (CHA) walker[3], which is switched on/off at pH 7.5 and 5.0, respectively. With this strategy, it is verified that on microparticle surface, the walking rate of the designed DNA walker was controlled in an efficient way by changing the pH gradients. Notably, we have also demonstrated that association and dissociation of the walker on microparticles could be dynamically controlled under alternate stimulation of acidic and alkaline pHs. This pH-dependent walker could be possibly applied as a pH-monitoring nanomachine in living cells and biohybrid micromotor for mimicking chemotactic motions of living organisms.

Reference

[1] Idili, A.; Vallée-Bélisle, A.; Ricci, F., *J. Am. Chem. Soc.* **2014**, 136, 5836.
[2] Leitner, D.; Schröder, W.; Weisz, K., *Biochemistry* **2000**, 39, 5886.
[3] Jung, C.; Allen, P. B.; Ellington, A. D., *Nat. Nanotechnol.* **2016**, 11, 157.

Data storage in DNA with fewer synthesis cycles using composite DNA letters

Authors: Leon Anavy[1, *], Inbal Vaknin[2], Orna Atar[2], Roee Amit[2], Zohar Yakhini[1, 3, *]

[1] Computer Science Department, Technion, Haifa, 3200000, Israel

[2] Faculty of Biotechnology and Food Engineering, Technion, Haifa, 3200000, Israel

[3] School of Computer Science, Herzliya Interdisciplinary Center, Herzliya, 4610101, Israel

*Correspondence to: anavy@technion.ac.il, zohar.yakhini@gmail.com

The density and long-term stability of DNA make it an appealing storage medium, particularly for long-term data archiving. Existing DNA storage technologies involve the synthesis and sequencing of multiple nominally identical molecules in parallel, resulting in information redundancy. We report the development of encoding and decoding methods exploiting this redundancy using composite DNA letters. A composite DNA letter is a representation of a position in a sequence that consists of a mixture of all four DNA nucleotides in a pre-determined ratio. Our methods encode data using fewer synthesis cycles. We encode 6.4MB into composite DNA, with distinguishable composition medians, using 20% fewer synthesis cycles per unit of data, compared with previous reports. We simulate encoding with larger composite alphabets, with distinguishable composition deciles, to show that 75% fewer synthesis cycles are potentially sufficient. We describe applicable error correcting codes, inference methods and investigate error patterns, in the context of composite DNA letters.

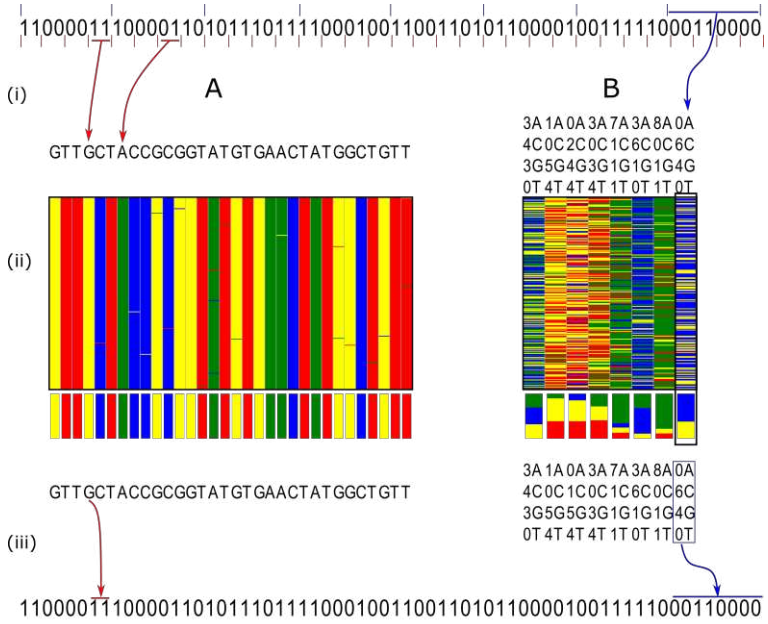


Figure 1: Encoding a binary message using standard and composite DNA. A binary message, depicted on top, is encoded into DNA. A. Standard DNA based storage scheme[7]. The binary message is being encoded to DNA by mapping every 2 bits (represented by the short red separating lines) to a DNA letter or synthesis cycle (i), the designed DNA sequence will then be synthesized and sequenced by a noisy procedure that introduces some errors (ii). The sequencing output is then used to infer the DNA composition at every position (iii). Decoding of the original message is done assuming the use of an error correcting code. B. The same message is encoded using a composite DNA alphabet of resolution $k = 10$ by mapping every 8 bits (represented by the blue separating lines) of the binary message to a single composite DNA position (a single synthesis cycle when using appropriate hardware). Sufficiently deep sequencing allows to correctly identify the original composite letters and to decode the message. The decoding also uses an error correction mechanism (Reed-Solomon over the appropriate finite field, in our implementation.)

Structure sampling for rate estimation in tethered molecular circuits

Matthew R. Lakin, Julian M. Weisburd, Sarika Kumar
Department of Computer Science and Center for Biomedical Engineering
University of New Mexico, Albuquerque, NM 87131, USA. mlakin@cs.unm.edu

Abstract: In order to implement systems capable of robust, rapid decision making, and to implement larger and more sophisticated decision-making algorithms, it can be advantageous to implement the computation using molecular components that are tethered to a DNA tile nanostructure. Attaching circuits to a surface accelerates computations by anchoring components in close proximity, and it also enhances scalability by permitting reuse of identical DNA sequences at different locations. These systems show great promise for speeding up, and scaling up, molecular computations and for practical applications including delivery of theranostic molecular circuits to cells. In the latter case, the DNA nanostructure itself could confer protection on the attached molecular logic circuit cargo.

A key goal in molecular computing research is to facilitate rapid and reliable design of molecular circuits, including tethered molecular circuits. The localization of components in tethered molecular reaction systems means that molecular geometry becomes important in determining whether nearby components may in fact interact, and at what rate. In previous work, we used a constraint-based approach to determine whether molecular components on a surface are in fact arranged such that they may actually interact. Here we attempt to probabilistically estimate the rate constants of localized molecular interactions using structure sampling based on a simple biophysical model. We show that even a very simple model can yield results with a factor of three of rate constants inferred from experimental data (Chatterjee et al, Nature Nanotechnology 12, 920-027, 2017). Our work will thus provide a basis for subsequent enhanced modeling, simulation, and design tools for tethered molecular circuits.

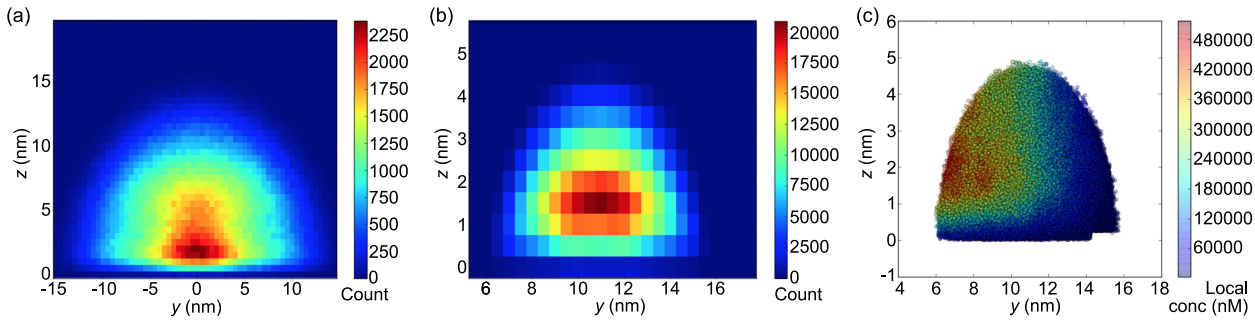


Figure: Results from preliminary experiments on *in silico* structure sampling and rate computations, using the worm-like chain model for single-stranded domains. (a) Heatmap showing the (2D) distribution of a reactive toehold in the system of Chatterjee et al, with fuel bound. (b) Heatmap showing the (2D) distribution of the reactive toehold in a second hairpin structure. (c) Scatter heatmap showing how the local concentration observed at the second toehold differs with its physical location.

Independent Control of the Thermodynamic and Kinetic Properties of Aptamer Switches

Brandon D. Wilson[1], Amani Hariri[2], Ian Thompson[2], and H. Tom Soh[2,3,4*]

[1]Department of Chemical Engineering, [2]Department of Electrical Engineering, [3]Department of Radiology, Stanford University, Stanford, CA 94305, USA.[4]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA.

Molecular switches that change their conformation upon target binding offer powerful capabilities for biotechnology and synthetic biology. In particular, aptamers have proven useful as switches because they offer excellent binding properties, undergo reversible folding, and can be readily engineered into a wide range of nanostructures. Unfortunately, the thermodynamic and kinetic properties of the aptamer switches developed to date are intrinsically coupled such that fast temporal resolution can only be achieved at the cost of lower sensitivity or increased background signal[1].

In this work, we describe a general design strategy of decoupling the thermodynamic and kinetic behavior of aptamer switches to achieve independent control of sensitivity and temporal resolution. We explore the use of an intramolecular strand-displacement (ISD) strategy[2] and the degree to which its binding properties can be controlled through rational design. In this construct, a molecular switch is generated by coupling an aptamer to a partially complementary displacement strand via a poly-T linker (**Figure 1**). This design offers two distinct control parameters: displacement strand length[3] (L_{DS}) and loop length[4] (L_{loop}). This is in contrast to alternative molecular switches such as aptamer beacons, which have just a single control parameter—displacement strand length—resulting in coupled kinetics and thermodynamics[1]. Using an ATP[4] aptamer as an exemplar, we provide the theoretical basis of our design and experimentally characterize an array of aptamer switches with effective dissociation constants (K_D) ranging from 10 μM to 40 mM and binding kinetics ranging from 170 ms to 3 s – all generated from a single parental aptamer[6]. Importantly, we are able to modulate the kinetics of the switching response while maintaining the same effective binding affinity (**Figure 2**). Our strategy should be broadly applicable to other aptamers, enabling the development of switches with characteristics suitable for diverse applications in biotechnology..

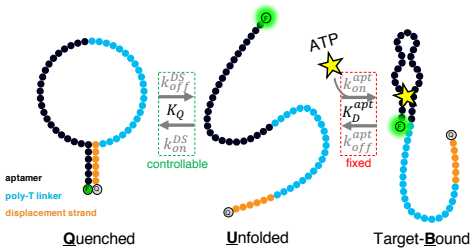


Figure 1 | Overview of intramolecular strand displacement and relevant parameters. We use intramolecular strand-displacement to convert an existing aptamer into a signaling switch. Target binding depletes the unfolded population, shifting the reaction to the right, generating a signal that is proportional to target concentration. Reducing L_{DS} increases effective binding affinity and increases background. Decreasing L_{loop} decreases background and decreases binding affinity. Decreasing either parameter will increase overall rate of binding, rendering it possible to increase the kinetics of molecular recognition while retaining the same K_D^{eff}.

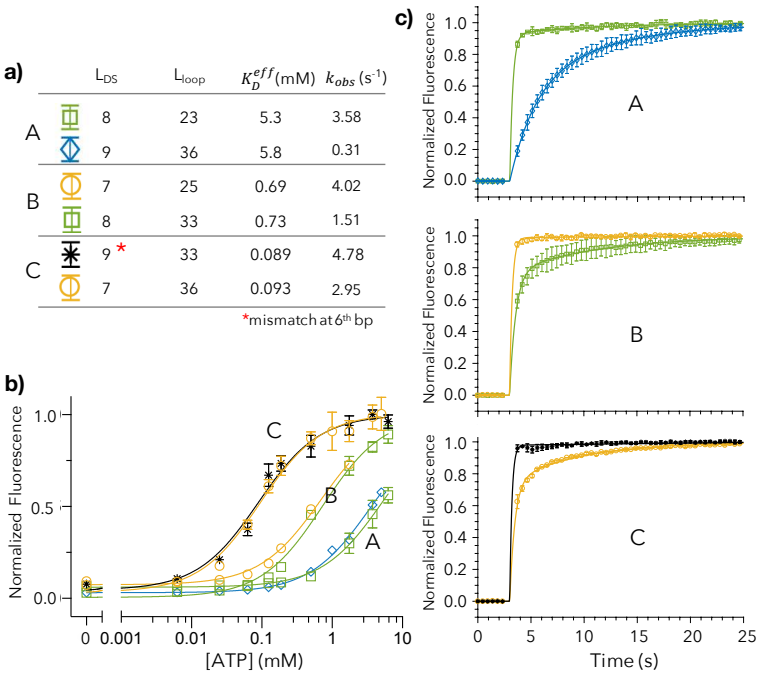


Figure 2 | Independent tuning of kinetics and thermodynamics through simultaneous changes to L_{DS} and L_{loop}. Three pairs of ISD constructs **(a)** exhibit nearly identical binding curves **(b)**, yet they have been engineered in terms of L_{DS} and L_{loop} to exhibit vastly different kinetic responses **(c)**.

(1) Hall, B.. *Biotechnol. Bioeng.* **2009**, *103* (6), 1049–1059.
(2) Tang, Z. JACS **2008**, 130, 34, 11268–11269.
(3) Vallee-Belisle, A. PNAS **2009**, 106 (33) 13802-13807
(4) Mariottini, D. JACS **2018**, 140, 44, 14725014734
(5) Huizenga, D. E. *Biochemistry* **1995**, *34* (2), 656–665.
(6) Zhang, Z. *Nucleic Acids Res.* **2017**, *45* (13), 7593–7601.

Development of a Single Probe Domain for Multiplexed Nucleotide Profile Recognition

Seung Won Shin[1], Charles Park[2,3], James J. Moon[2,3,4], Dan Luo[5], Soong Ho Um[1,6]

[1]School of Chemical Engineering, Sungkyunkwan University, Suwon, Gyeonggi-do, South Korea
[2]Biointerfaces Institute, University of Michigan, Ann Arbor, MI 48109, USA
[3]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[4]Department of Pharmaceutical Sciences, University of Michigan, Ann Arbor, MI 48109, USA
[5]Department of Biological and Environmental Engineering, Cornell University, Ithaca, NY 14850, USA
[6]SKKU Advanced Institute of Nanotechnology (SAINT), Sungkyunkwan University, Suwon, Gyeonggi-do, 16419, South Korea

In living organisms, the sequence of nucleotides enables the fine-tuning of biological activities and also conveys essential heritage information to descendants. As is well-known, one of the most important features of nucleotides is their ability to interact with each other in a complementary manner, a process called hybridization. The hybridization of nucleotides in perfect complementation is not only stable but highly selective. This complex and delicate reaction can be interpreted with simple thermodynamic principles. The parameters to calculate the thermodynamic properties of the nucleotides involve enthalpy, entropy, and Gibb's free energy. For instance, the melting temperature (Tm) of nucleotides is precisely described by the Van't Hoff equation, and the hybridization interaction between nucleotides is well-described by the nearest-neighbor model in a sequence-specific manner. Moreover, computational tools have enabled prediction of detailed kinetics of hybridization and have contributed to development of nucleic acid engineering.

If there are a few mismatches in a complementary sequence, hybridization becomes weaker and less stable. Even though the hybridization is relatively weak, the presence of mismatches in nucleotides often provides the potential to fine-control reactivity. In the case of RNA interference, when the interfering RNA fragment, such as siRNA or microRNA, is the perfect complement of the target region of mRNA, hybridization at the complementary site leads to endonucleolytic cleavage and degradation of mRNA. However, when there are mismatches, the complex forms a bulged structure. This mismatched bulge prevents the complex from degrading but makes it possible to delicately regulate the efficiency of protein expression. A group of microRNAs sharing sequence similarity is categorized as a microRNA family. Their sequences are diversified by only a few bases, but they manipulate thousands of mRNAs in a cooperative and harmonized manner. Thus, the diversity of sequences effectively enhances the functionality of a nucleotide. In other words, the information is spread out for better functionality.

What we should consider is that the spreading out of information might be better for functionality but also obscures the information itself. For a better understanding of and to unearth the motivation of the metabolic network, we need to reduce diversity and concentrate on the information inherited in the groups of nucleotide analogs. In this study, we tried to reveal the existence of a representative sequence among multiple oligonucleotides through simple calculations of strand-to-strand interaction. As study models, 8 to 10-based random nucleotides and their analogs were used as an 'origin' and its 'mutants'. We assumed that the mutants are the facets of the diversified form of the origin for better functionalization. With sufficient similarity, both the origin and the mutants would share the same purposes to reshape the metabolic networks. If it is possible to create a single representative sequence from multiple mutants, the analogs of the origin, it would be possible to reveal inherited biological purposes.

Keywords: representative nucleotide, hybridization profile, multiple equilibria, sociogram

A Computational Framework for DNA Sequencing-Based Microscopy

Ian T. Hoffecker[1], Yunshi Yang[1] , Giulio Bernardinelli[1], Pekka Orponen[2], and Björn Högberg[1]

Presenter
[1] Dept. of Medical Biochemistry and Biophysics, Karolinska Institutet, S-17177 Stockholm, Sweden, http://hogberglab.net
Correspondence to bjorn.hogberg@ki.se and ian.hoffecker@ki.se
[2] Department of Computer Science, Aalto University, 00076 Aalto, Finland

Keywords

DNA computing · topology · next-gen sequencing · graph theory · polonies ·

Abstract

DNA has been explored as an information storage and transmission medium for various artificial applications, and herein we describe a concept for a DNA sequencing-based microscopy method. We conducted an *in silico* proof of concept of microscopic image construction based on 2D networks of barcoded locally-confined solid phase polymerase chain reactions (surfaces polonies) whose identities are recovered through next gen sequencing as opposed to optical means. Information is stored as topology or connectivity of adjacent polonies that, when grown to saturation form concatemers – imparting pairwise association between spatially neighboring patches. Through judicious use of barcoding, polonies become a kind of pixel, the complete network of which forms a spatial map of relative locations. In our implementation, we simulate a polony network and form a set of pairwise associations that are then scrambled, removing any record of their locations. We then show reconstruction of the network by planar embedding and three algorithms (a non-deterministic spring relaxation, one based on the Tutte embedding, and another based on coordinate transformation of a topological distance matrix) is sufficient to reconstruct the original positions of the network with high global fidelity. We also demonstrate a concept for how network reconstruction could be extended for image formation through the association of additional molecules with readable tags that would be associated by contact with the underlying polony network and sequenced alongside the pairwise association data. A three target scheme representing artificial colors can then be used to convert the reconstructed polony network into an array of pixels colored according to their composition of associated targets. We examine the distortions of such reconstructions and show that for independently seeded networks, distortions are not long range, and primarily persist at the scale of individual polonies or boundaries of the region of interest. We hold that this framework and pipeline for reconstruction could be exploited for image analysis of micro- and nano-scale surfaces with molecular libraries of potentially high multiplicity and with the throughput offered by next gen sequencing and automated reconstruction algorithms.

Live dynamical tracking of mRNA transcripts in single *E. coli* cells reveals bursts of fluorescence degradation

Naor Granik, Noa Katz, Yoav Shechtman, and Roee Amit

The process of gene expression is comprised of many successive stochastic events, all of which depend on complex interactions between a variety of cellular components (transcription factors, polymerases, etc.). The outcome of this is a phenotypic variability between cells which can have significant physiological consequences in natural and synthetic systems (1).
One noteworthy source of noise in gene expression is fluctuations in promoter activity (2). Promoter architecture, defined by the type of polymerase involved and the characteristics of transcription factor binding sites, has a strong effect on the level of activity and transcriptional noise. However, the precise relationship between this architecture and transcriptional dynamics has not yet been established.
At present, research into transcriptional dynamics is largely based on two complementary imaging approaches: SM-FISH measurements which cannot directly probe temporal effects (3), and live dynamical tracking of large RNA cassettes that are typically composed of 24 binding sites for RNA binding protein-fluorescent protein (RBP-FP) fusions (4). However, live tracking of single molecules in single cells is often not possible in many applications due to limitations of the RBP binding cassette. These are mostly due to cassette size and fluctuations in RBP occupancy, which together camouflage successive transcriptional events (5).
Using new RBP-FP cassettes designed in our lab we track mRNA molecules tagged with cassettes containing only four binding sites for PP7 or Qβ coat proteins. These short cassettes are stably bound and allow us to separate successive transcriptional events at high resolution. We implemented these cassettes by installing them at the transcriptional start site of a T7-promoter in *E. coli*.
Our results reveal that in steady-state, T7 transcription forms fluorescent puncta composed of ~15-30 mRNA molecules at the cell's poles, which exhibit a multi-state random telegraph noise (RTN) consisting of both positive and negative fluorescent bursts, spaced by periods of non-classified or quiescent activity. We identify positive events with the previously observed transcriptional bursts, and term the negative events fluorescent signal degradation bursts. We show that both signal degradation and transcriptional bursts consist typically of 1-5 transcripts, and that the duration of the non-classified states behaves according to an exponential distribution characterized by a 10 min decay time.
We support the experimental findings by showing that both a multi-state RTN simulated signal, and Monte Carlo simulations based on a kinetic model which allows for two possible degradation rates provide an adequate description to our data. Furthermore, from numerical simulations we conclude that our extended model leads to a reduced Fano Factor as compared with the two-state model. As a result, we conclude that "bursts" may be a significantly more pervasive phenomenon than previously thought, and that our observations reflect the underlying biophysical structure of the observed puncta. This, in turn, has implications to not only transcriptional dynamics, but also to emergent biophysical properties of membrane-free compartments that have gained increased attention in recent years.

References:

1. Elowitz, M.B. 2002. Stochastic Gene Expression in a Single Cell. Science. 297: 1183–1186.

2. Blake, W.J., G. Balázsi, M.A. Kohanski, F.J. Isaacs, K.F. Murphy, Y. Kuang, C.R. Cantor, D.R. Walt, and J.J. Collins. 2006. Phenotypic Consequences of Promoter-Mediated Transcriptional Noise. Molecular Cell. 24: 853–865.

3. Raj, A., P. van den Bogaard, S.A. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods. 5: 877–879.

4. Tutucci, E., M. Vera, J. Biswas, J. Garcia, R. Parker, and R.H. Singer. 2018. An improved MS2 system for accurate reporting of the mRNA life cycle. Nature Methods. 15: 81–89.

5. Jones, D., and J. Elf. 2018. Bursting onto the scene? Exploring stochastic mRNA production in bacteria. Current Opinion in Microbiology. 45: 124–130.

A Universal Self-Replicating Computer using DNA Sticker CAM

Mark G. Arnold and Israel M. Martinez-Perez

In the beginning, the Roweis-Winfree sticker system proposed using a conventional computer to control tubes containing DNA strands that process bits denoted with stickers (codewords ≈ 10 bases). Massive SIMD parallelism was to compensate for the slowness and unreliability of DNA as a computing medium, but no practical wetlab implementations exist. In a different context, stickers were extended with staples to create DNA nanostructures programatically. Finally, Sarkar and Ghosal offered a new sticker variant using SISD (one strand per tube) instead of SIMD (trillions of strands per tube), but this is impractical because DNA is slower than the conventional computer, which could implement the entire SISD system more easily. We imagine a further thought experiment where the sticker/staple system's electronic computer would be replaced with a biochemical computer, whose tubes are fabricated as DNA nanostructures held together by hybridization using codewords that are unique and non-interfering with the codewords that implement the sticker data processed by this physical nanostructure. Its fabrication is a by-product of running software on its mother sticker/staple computer, requiring unique codewords for several (≥ 3) generations. Such a self-replicating system has problems similar to living organisms: What does it eat? Where does waste go? We imagine a lab-on-chip "super-mother" MEMS could feed it single-stranded DNA and dispose of waste created in algorithms that need to forget old values. The super-mother MEMS is a silicon-based sticker/staple computer that creates the first generation nanostructure mother using the same replication software which that mother and her offspring will use. Each generation will copy (the same bits of) this software from its codewords into its daughter's codewords. This poster considers one element of this thought experiment: how a DNA-nanostructure can execute conventional software algorithms from a DNA-based memory. This suggests the need for a universal sticker machine, whose simple control state machine remains fixed, but which can solve any computable problem by reprogramming machine language represented in a DNA sticker memory. Although the interpreter of this machine language follows the Sarker-Ghosal model, the program being interpreted is free to process with the more capable Roweis-Winfree SIMD model. Conventional processors use (non-content addressable) RAM and have two distinct registers: a program counter (PC) and an Instruction Register (IR). The proposed processor needs a Content Addressable Memory (CAM), where contents and addresses are both recorded on each strand. The Sarkar-Ghosal model maps each register to a tube; here, PC and IR merge into a PCIR tube, which holds a single strand with both an instruction and its associated address. Control flow is implemented by choosing one instruction/address strand from many in a Memory (M) tube and moving that strand to PCIR (after transfering the old strand from PCIR back to M). During normal execution, no new machine-language strands are created, copied or destroyed; they simply cycle in and out of PCIR following the flow of control of the program. Although it would not be the normal control flow, the system could be designed (at the cost of hybridization) to allow self-modifying programs, and therefore be universal in the fullest sense. Conventional architectural tradeoffs, like instruction formats and multi-cycle-versus-pipelined datapaths, are applicable, with the focus here on the simplest multi-cycle, one- or two-address formats. As in earlier sticker systems, data manipulation requires hybridization to set bits and attach staples. Surprisingly, most control flow, including branch and subprogram, occurs without setting any bits, offering substantial efficiency compared to a machine-language interpreter implemented naively with stickers.

Efficient size estimation and impossibility of termination in uniform dense population protocols

David Doty[1]
Department of Computer Science, University of California, Davis
doty@ucdavis.edu

Mahsa Eftekhari[2]
Department of Computer Science, University of California, Davis
mhseftekhari@ucdavis.edu

Abstract

We study *uniform* population protocols: networks of anonymous agents whose pairwise interactions are chosen at random, where each agent uses an *identical* transition algorithm that does not depend on the population size n. Many existing polylog(n) time protocols for leader election and majority computation are nonuniform: to operate correctly, they require all agents to be initialized with an approximate estimate of n (specifically, the value $\lfloor \log n \rfloor$).

Our first main result is a uniform protocol for calculating $\log(n) \pm O(1)$ with high probability in $O(\log^2 n)$ time and $O(\log^4 n)$ states ($O(\log \log n)$ bits of memory). The protocol is not *terminating*: it does not signal when the estimate is close to the true value of $\log n$. If it could be made terminating with high probability, this would allow composition with protocols requiring a size estimate initially. We do show how our main protocol can be indirectly composed with others in a simple and elegant way, based on *leaderless phase clocks*, demonstrating that those protocols can in fact be made uniform.

However, our second main result implies that the protocol *cannot* be made terminating, a consequence of a much stronger result: a uniform protocol for *any* task requiring more than constant time cannot be terminating even with probability bounded above 0, if infinitely many initial configurations are *dense*: any state present initially occupies $\Omega(n)$ agents. (In particular no leader is allowed.) Crucially, the result holds no matter the memory or time permitted.

Finally, we show that *with* an initial leader, our size-estimation protocol can be made terminating with high probability, with the same asymptotic time and space bounds.

[1] Supported by NSF grants 1619343 and 1844976.
[2] Supported by NSF grants 1619343 and 1844976.

General-purpose analysis package for coarse-grained simulations of DNA/RNA nanotechnology

Erik Poppleton[1], Michael Matthies[1], Shuchi Sharma[1], Joakim Bohlin[2], Petr Sulc[1]
[1]The Biodesign Institute, Arizona State University
[2]Physics Department, University of Oxford, Oxford, UK

We present here a package of design, visualization and analysis tools that extend the functionality of the oxDNA/oxRNA [1,2] coarse-grained models for DNA/RNA simulations. These models have frequently been used as tools to investigate the biophysics of nanostructure assembly and dynamics and to investigate properties of assembled nanostructures. However, previous studies using these models have performed their analysis in a manner unique to their individual system and relied on custom analysis tools developed specifically for each project. While these studies have produced informative results, generalizing their methods to other designs is non-trivial proposal. This limits the usage of molecular modeling to groups with the skill set to design and program structure analysis pipelines. We present a package of visualization, analysis, and clustering modules that are applicable across a wide-range of designs and we hope will improve the accessibility of molecular simulations to the DNA/RNA nanotechnology community. We demonstrate examples of applications across the size-scales of nanostructures, from single DNA junctions, to small RNA tiles, to 2D and 3D DNA origamis.

The core of this package is a new visualization tool, oxDNA-viewer [3], a web browser-based app that natively reads oxDNA files. We intend to eventually make it an available on a web-server with limited simulation capacity. Unlike previous visualizers for oxDNA files, oxDNA-viewer dynamically reads trajectory files, which allows you to load and create movies from multi-gigabyte trajectories of large origamis. It also has some editing capabilities, allowing users to simulate individual components of larger assemblies separately, bring them together in a graphical interface, and export the combined structure for further simulation. This is especially useful for structures exported from CaDNAno and Tiamat, which are often drawn as modular units. OxDNA-viewer also provides a platform for data overlays representing parameters extracted from simulations.

These analysis tools for oxDNA/oxRNA further include a set of command-line based tools for automated simulation evaluation. Each tool takes an oxDNA trajectory file and produces a data output which can then be visualized in oxDNA-viewer. Examples of analysis performed include angles between duplexes in nanostructures, residence of hydrogen bonds, mean structure and RMSD, and unsupervised clustering into subpopulations corresponding to distinct conformations adopted during simulation (Fig 1). We emphasize that these scripts are not tailored for a specific design, and the same scripts were used on the examples presented.

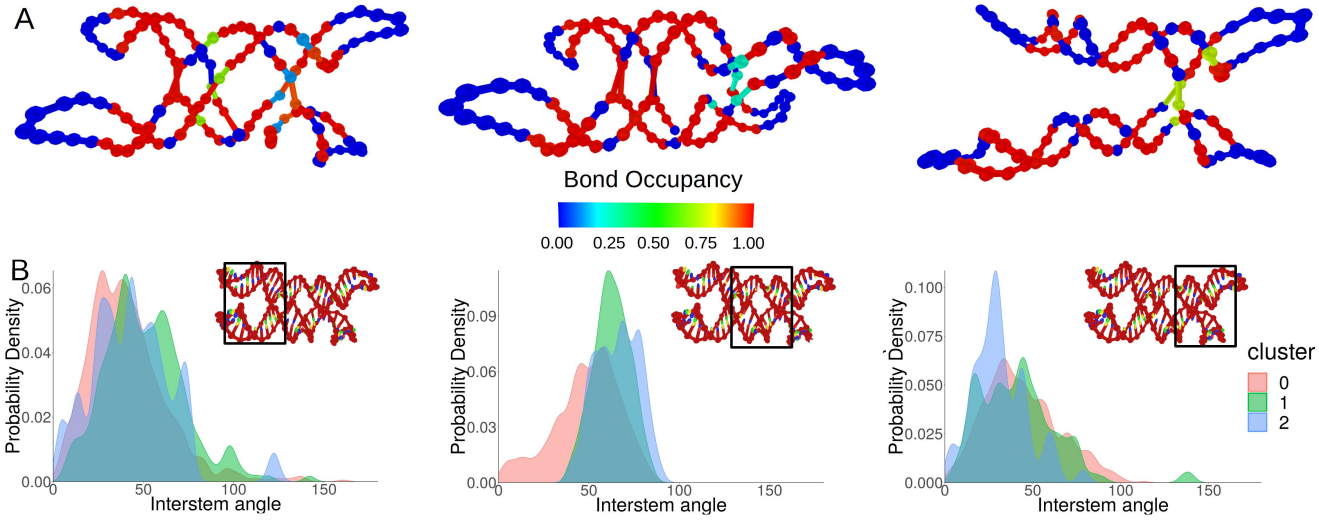


Fig 1. (A) A simulation trajectory of a single-stranded RNA tile was clustered into three structurally distinct groups using unsupervised machine learning. The mean positions of the nucleotide centers of mass for each cluster were then reconstructed using multidimensional scaling and the hydrogen bond occupancy was compared with the original design file. The mean structure was then overlaid with the bond residency using oxDNA-viewer. (B) Analysis of the angles between each of the three parallel duplex pairs (inset) were compared between the three clusters, revealing futher distinct structural features.

[1] B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, J. P. K. Doye. Introducing Improved Structural Properties and Salt Dependence into a Coarse-Grained Model of DNA . *J. Chem. Phys.* **142**, 234901 (2015).
[2] P. Šulc, F. Romano, T. E. Ouldridge, J. P. K. Doye, A. A. Louis. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* **140**, 235102 (2014).
[3] https://github.com/sulcgroup/oxdna-viewer

Transduction and Processing of Protease Activity Using DNA

Hieu Bui[1,2], Mario G. Ancona[2], Ellen R. Goldman[2], Yongchan Kim[2], Divita Mathur[2,3], Igor Medintz[2]*

[1]National Research Council, Washington, D.C. USA.[2]Naval Research Laboratory, Washington D.C., USA.[3]George Mason University, Fairfax, Virginia, USA.*Email: hieu.bui.ctr@nrl.navy.mil

Abstract: Proteins are a primary material of all living cells, and so the ability to break them down by hydrolyzing their peptide bonds is a critical biological function accomplished by enzymes called proteases. Given their ubiquity and importance, there is tremendous value for science and medicine in being able to monitor and analyze protease activity with high specificity and spatiotemporal resolution. With this as motivation, we here describe the development of a novel nanosystem that can translate enzymatic activity into DNA signals that are then readily amplified, combined, processed, and output. Unlike traditional biosensing methods (e.g., immune-PCR, or peptide-antibody-DNA conjugation), our system employs DNA nanotechnology to realize a flexible and modular implementation [1]. In particular, the design consists of two parts as illustrated in Figs. 1A and 1B, one an ordinary DNA hairpin and the other a hybrid loop consisting of a DNA oligo conjugated with a peptide target. The two parts share complementary DNA segments but do not hybridize because of mechanical and topological constraints unless the specific protease is present to sever the loop. The opening of the loop provides the conformational freedom needed for the loop DNA to engage with and open the hairpin, which in turn exposes a DNA sequence that represents the output signal of the transducer. This functionality is demonstrated with both electrophoresis and fluorescence in Fig. 1C. These data also exhibit the potential generality of the approach in that we created a second sensor system specific to a different protease simply by swapping out the target peptide. Furthermore, the two sensors can be combined to perform a logical OR function as shown in Fig. 1D. We regard this as a particular advantage of DNA in being able to perform downstream information processing, which is especially valuable for combatting false positives. More broadly, this system can be generalized beyond protease detection to focus on nucleases [2] as well as other biomarkers, and thus may represent a general platform for biosensing that could potentially be used *in vivo* as well.

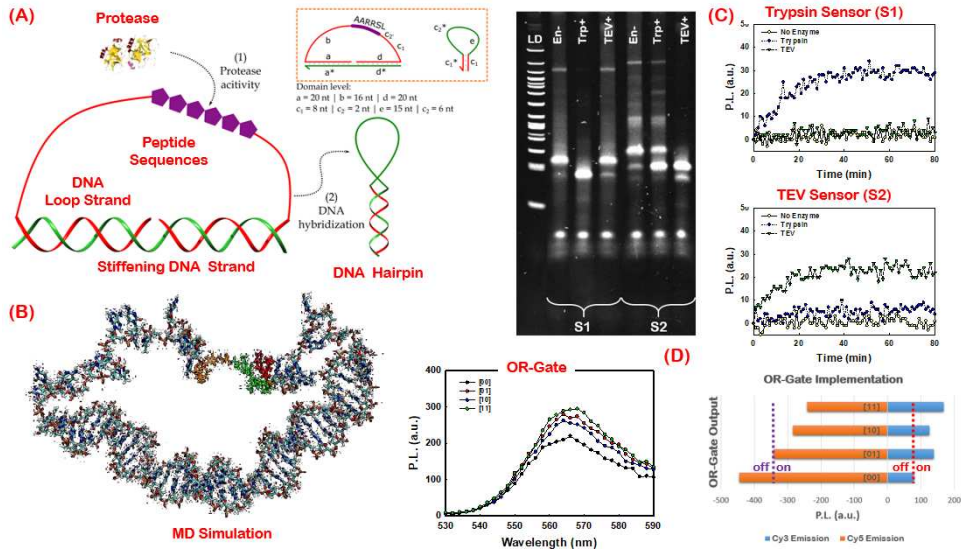


Figure 1: (A) The modular design of the loop sensor for proteases. (B) Molecular dynamic simulation validating the structural integrity. (C) Experimental demonstration of detecting individual protease. (D) Experimental implementation OR-gate by sensing either or both proteases.

Acknowledgement: This work was supported by ONR's Biomaterials and Bionanotechnology Program (Dr. Laura Kienker, Code 34) and NRL's Nanoscience Institute.

[1] Bui, H., *et al.,* 2019. *Small*, 15, 1805384. [2] Bui, H. *et al.* 2019, FNANO.

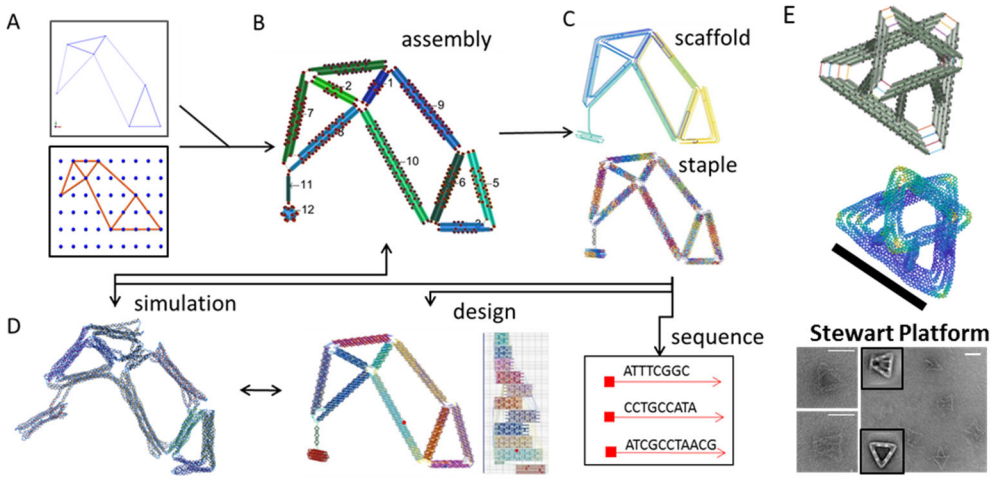
Design automation for DNA origami mechanisms

Chao-Min Huang[1], Anjelica Kucinic[2], Carlos E. Castro[1, 3] and Hai-Jun Su[1*]*

*Corresponding authors: castro.39@osu.edu su.298@osu.edu

DNA nanotechnology and scaffolded DNA origami take advantage of programmed Watson-Crick base pairing to construct self-assembled nanostructures. DNA origami mechanisms coordinate the rigidity of dsDNA and the flexibility of ssDNA to create dynamic structures with desired motions[1]. While designing DNA origami structures can be challenging, recent efforts have led to automated approaches for design[2]; however, these design tools focus on static objects and the rigidity of the components limits their usage in dynamic mechanism designs. Current design processes for DNA origami mechanisms and machines still rely on heavily manual DNA routing approach within caDNAno software[3] that can be highly challenging for multi-component structures. Hence, designing complex mechanisms requires prior expertise or training, and although dynamic mechanisms have a range of applications, their broader development is significantly limited by the challenges of design.

We present a GUI-based design software in which users are able to insert bundles as components with various lengths, cross-sections, and underlying cross-section lattices. The software allows for easy design of multiple components and translation/rotation of individual components to approximate positions in 3D space. Then, the software automatically (or manually for expert users) specifies connection points between components for flexible or rigid joints. Later, a graph-theory algorithm is used to generate the overall scaffold routing and staple routing. Finally, this software is able to export the staple sequences for fabrication and experiments, compatible JSON files for design post-processing or fine modification in caDNAno to the software, and oxDNA files for simulation[4]. These features ultimately facilitate optimizing the design through iterations between designs and simulation to provide a robust design framework. We envision that this software promotes structual complexity in DNA origami and further benefits related fields and appications in DNA nanotechnolgy.



Reference:

[1]. Marras, et al. *PNAS U.S.A.* 2015

[2] Veneziano, et al. *Science* 2016

[3] Douglas, et al. *Nucleic Acids Res.* 2009

[4] Šulc, et al. J. Chem. Phys. 2012

Figure 1 Design process of this software. (A)Start with STEP files from other CAD software or the embedded sketch tool. (B) Assemble 3D components in a graphical user interface. (C) Automate the scaffold and staple routing. (D) Result outputs: (Left) oxDNA topology and configuration files. (Middle) caDNAno JSON file and its analogue in the software with 3D mapping for modifying. (Right) staple sequences for experiment (E) Design, simulation and experiment for the Stewart platform. Scale bar =50 nm.

Availability-Driven Design of Hairpin Fuels and Small Interfering Strands for Leakage Reduction in Autocatalytic Networks

Drew Lysne[1], Kailee Jones[1], Alma Stosius[1], Timothy Hachigian[1], Jeunghoon Lee[1,2] and Elton Graugnard[1]
[1]Micron School of Materials Science and Engineering, [2]Department of Chemistry and Biochemistry, Boise State University, 1910 University Dr., Boise, Idaho 83725, United States

Leakage, producing an output in the absence of an input, limits the utility of DNA-based circuits and computational tools for biomedical and technological applications [1]. Strategies that have been explored to suppress or eliminate leakage include DNA clamps [2], mismatching [3], locked nucleic acids (LNA) [4], availability and mutual availability [5], domain level redundancy [6], rigorously purified strands [7], multi-arm junctions [8], and shadow circuits [9]. As a new tool for leakage reduction, this study utilizes availability to reduce leakage in an entropy-driven autocatalytic DNA reaction network [10]. Here, we report the direct tailoring of fuel strand availability through two novel approaches: (1) the addition of interfering domains to fuel strands, and (2) the introduction of separate small interfering strands. In each case, these approaches were designed to reduce the availability for key domains within fuel strands. The strands were designed against available regions of the substrate complex analyzed by NUPACK availability calculations. The modified fuels showed decreased leakage but also decreased catalytic activity. Successful fuel designs resulted in increased performance ratios, defined as $t_{1/2\ leak}/t_{1/2\ 0.1x\ catalyst}$, ($t_{1/2}$ refers to time to half completion) of up to 22%. Employing the small interfering strand (5-12 nucleotides) improved the performance ratios by up to 21%. Furthermore, the stability of the network using both leakage reduction methods were highly predictable using computed availability, which accounts for the availabilities of both the fuel and substrate backbone (**Figure 1**). Computed availability and experimental results show a Pearson's correlation coefficient of -0.86 for modified fuel strands and -0.89 for fuel and interfering strand combinations.

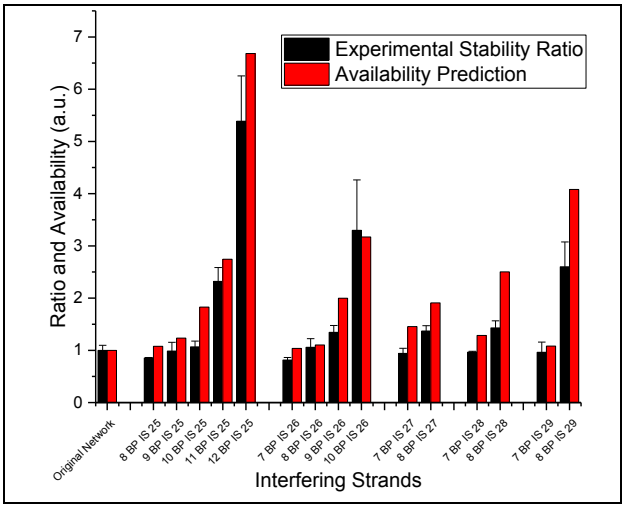


Figure 1. Availability (red) compared to stability ratio (black) for interfering strands. Availability was calculated using the formula: $[1/\sum_{24}^{28}(P_{F(i)}P_{B(i^*)})]$, where $P_{F(i)}$ and $P_{B(i^*)}$ are the availabilities of bases i and i^* of the fuel and substrate backbone respectively, and then normalized. The stability ratio was calculated as: $t_{1/2\ leak\ experimental\ fuel}/\ t_{1/2\ leak\ fuel}$.

The results of this study support the use of availability as a design metric for reducing leakage, but more work is needed to mitigate against the resulting catalytic losses. Thus, availability analysis can provide insight into regions of network vulnerability, which may be addressed through network design, domain, and sequence engineering.

1. Simmel FC, Yurke B, Singh HR (2019). Chem Rev 119:6326–6369. https://doi.org/10.1021/acs.chemrev.8b00580
2. Seelig G, Yurke B, Winfree E (2006). J Am Chem Soc 128:12211–12220. https://doi.org/10.1021/ja0635635
3. Jiang YS, Bhadra S, Li B, Ellington AD (2014).Angew Chemie Int Ed 53:1845-1848. https://doi.org/10.1002/anie.201307418
4. Olson X, Kotani S, Yurke B, et al (2017). J Phys Chem B 121:2594–2602. https://doi.org/10.1021/acs.jpcb.7b01198
5. Olson X, Kotani S, Padilla JE, et al (2017). ACS Synth Biol 6:84–93. https://doi.org/10.1021/acssynbio.5b00231
6. Thachuk C, Winfree E, Soloveichik D (2015). In: Proceedings of the 21st International Conference on DNA Computing and Molecular Programming - Volume 9211. Springer-Verlag New York, Inc., pp 133–153
7. Chen X, Briggs N, McLain JR, Ellington AD (2013).ProcNatlAcadSci 110:5386 LP5391.https://doi.org/10.1073/pnas.1222807110
8. Kotani S, Hughes WL (2017). J Am Chem Soc 139:6363–6368. https://doi.org/10.1021/jacs.7b00530
9. Song T, Gopalkrishnan N, Eshra A, et al (2018). ACS Nano 12:11689–11697. https://doi.org/10.1021/acsnano.8b07394
10. Zhang DY, Turberfield AJ, Yurke B, Winfree E (2007). Science (80-) 318:1121–1125

Molecular tagging with nanopore-orthogonal DNA strands

Katie Doroschak[1], Karen Zhang[1], Aishwarya Mandyam[1], Melissa Queen[1], Karin Strauss[1,2], Luis Ceze[1], and Jeff Nivala[1]

[1] Allen School of Computer Science and Engineering, University of Washington, Seattle, WA

[2] Microsoft, Seattle, WA

Molecular tagging is an approach to securely labeling physical objects using DNA or other molecules. In contrast to DNA data storage, molecular tagging systems are low-density information systems, analogous to RFID tags rather than hard drives. An ideal system should be inexpensive, quick and reliable to decode, and require minimal equipment. We created a molecular tagging system using DNA-based tags and the MinION nanopore device. Digital bits are represented by the presence or absence of distinct, nanopore-orthogonal DNA strands, called molecular bits (molbits). A single molbit is a DNA strand that modularly combines a unique barcode structure with a specific strand length. This enables us to identify molbits directly from the raw nanopore signal, significantly reducing the time required to decode a molecular tag and increasing tag accuracy. Similar to electronic message transmissions systems, error correction is added to resolve any remaining decoding errors in the molecular tag. Molbits are prepared for readout at the time of tag assembly and are then stabilized by dehydration. These steps extend the shelf life of the tag, decrease decoding time, and make it robust to contamination from environmental DNA. The result is an extensible, real time tagging system with high accuracy.

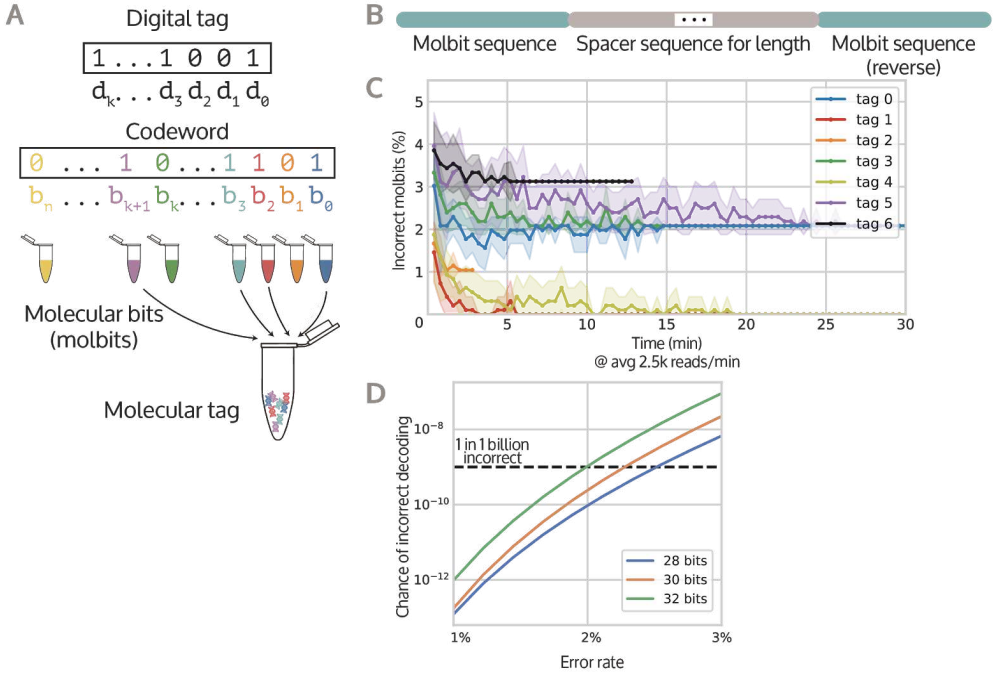


Figure 1. A) Molecular tag encoding scheme. (k=32, n=90) B) DNA strand composition for a single molecular bit. C) Molecular bit error rates per tag as a function of sequencing time, before error correction. Shorter run times were simulated via random sampling (n=50). D) Chance of whole tag identification error for varying message size and molecular bit error rate.

Composable computation in discrete chemical reaction networks*

Eric E. Severson[1], David Haley[1], and David Doty[1]

[1]University of California, Davis. {eseverson,drhaley,doty}@ucdavis.edu.

Abstract

We study the composability of discrete chemical reaction networks (CRNs) that *stably compute* (i.e., with probability 0 of error) integer-valued functions $f : \mathbb{N}^d \to \mathbb{N}$. We consider *output-oblivious* CRNs in which the output species is never a reactant (input) to any reaction. The class of output-oblivious CRNs is fundamental, appearing in earlier studies of CRN computation, because it is precisely the class of CRNs that can be composed by simply renaming the output of the upstream CRN to match the input of the downstream CRN.

Our main theorem precisely characterizes the functions f stably computable by output-oblivious CRNs with an initial leader. The key necessary condition is that for sufficiently large inputs, f is the minimum of a finite number of nondecreasing *quilt-affine* functions. (An affine function is linear with a constant offset; a *quilt-affine* function is linear with a periodic offset).

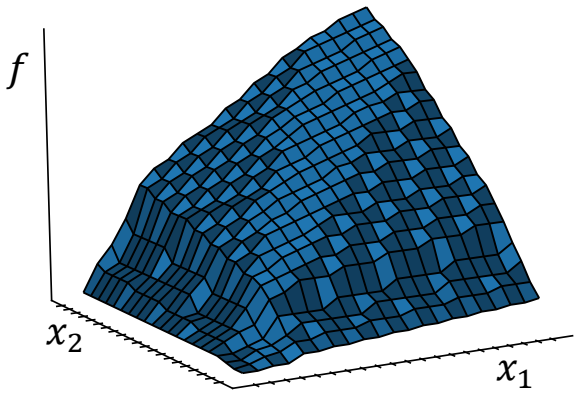


Figure 1: A representative example $f : \mathbb{N}^2 \to \mathbb{N}$ that is satisfies our main theorem, so is stably computable by an output-oblivious CRN.

*Authors supported by NSF grants 1619343 and 1844976.

Nanopore readout for scalable DNA circuit reporting

Karen Zhang[†1], Yuan-Jyue Chen[†2], Katie Doroschak[1], Karin Strauss[1,2], Luis Ceze[1], and Jeff Nivala[*1]
[1]*Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.*
[2]*Microsoft Research, Redmond, WA, USA. †equal contribution. *correspondence to jmdn@uw.edu.*

Strand displacement circuits are a popular strategy for DNA computing. Their enzyme-free nature and programmability give them a myriad of potential applications, from disease diagnostics to DNA-based artificial neural networks. However, traditional fluorescence-based readout of these circuits poses an obstacle to their versatility and scalability. Due to spectral overlap of traditional reporters (e.g. FAM, TAMRA, Cy5), optical detection of outputs is neither highly scalable nor multiplexable. Here, we present an alternate reporter method for real-time readout of toehold-mediated strand displacement DNA circuits using nanopore sensor technology. Our method enables simultaneous detection and quantification of multiple circuit elements in a single reaction volume. First, we demonstrate dynamic sensing of an operating circuit within the flow cell of a commercially-available high-throughput nanopore sensor array (Oxford Nanopore Technologies' MinION device). We then show that strand capture frequency can be correlated to concentration, allowing for direct quantification of desired circuit elements, enabling kinetic analysis. Finally, to investigate this reporter strategy's multiplexing potential, we present a collection of seven orthogonal circuit output sequences (strand barcodes) that can be classified with high accuracy at the single-molecule level from raw nanopore signal data using machine learning, with the potential to scale to much larger barcode sets. We conclude that nanopore-based detection of strand displacement circuits holds key advantages over fluorescence-based methods for real-time, multiplexed circuit readout on an inexpensive, portable sensor device.

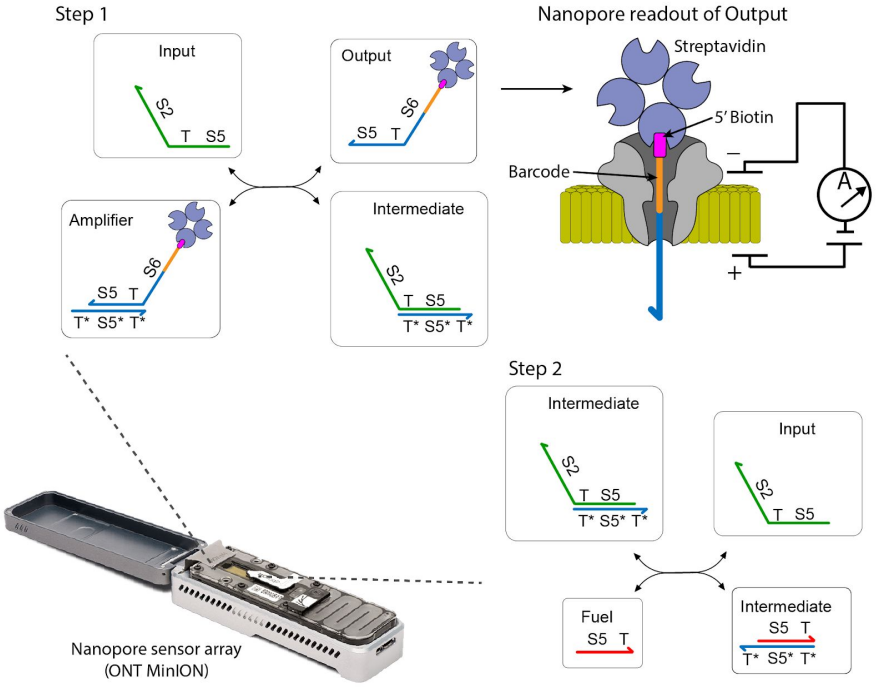


Figure 1. Catalytic seesaw gate circuit output detection with a nanopore sensor. Circuit components are mixed and loaded into a nanopore sensor array for real-time readout. Input strands react with the amplifier, which displaces the 5' labeled biotin-streptavidin output ssDNA (Step 1), which is then free to be captured and read by a nanopore sensor. The input strand is recycled (Step 2).

SOLQC: Synthetic Oligo Library Quality Control Tool

Omer Sabary*[a] , Yoav Orlev*[b], Roy Shafir[b], Leon Anavy[a], Eitan Yaakobi[a], and Zohar Yakhini[a,b]
* The two first authors contributed equally to this work
[a] Dept. of Computer Science, Technion. [b] Dept. of Computer Science, IDC Herzliya.

Recent years have seen a growing number and a broad scope of utilization of synthetic oligo libraries for various applications in synthetic biology. One important application of synthetic DNA libraries is **DNA storage**, where binary data is encoded as DNA molecules. This encoding requires a dedicated design of a synthetic DNA library. The designed molecules are synthesized, then sequenced and decoded to infer the encoded binary data. A second important application is **MPRAs - Massively Parallel Reporter Assays**, wherein a set of variants is examined with respect to a measurable property to support the better understanding of the sequence determinants of said property.

As experiments are growing by numbers and complexity, better analysis tools are necessary to facilitate quality control and to validate the results. Moreover, understanding the errors involved in the synthesis and the sequencing processes is important not only to control the quality of DNA libraries, but also to design appropriate error-correction codes and coding schemes when DNA is used for storage applications.

We present a novel analysis tool, called **SOLQC** tool, which enables fast and comprehensive analysis of synthetic oligo libraries. The SOLQC tool provides statistical information and insights about the quality and state of the library, including extensive graphical representation. The input to SOLQC is the result of running a sequencing reaction on a synthesized library. It comprises the designed variants and all sequenced reads. These reads can be filtered such that only valid reads will be processed. The selection of valid reads can be configured by the user according to their lengths and their barcodes (if such are used). SOLQC first matches between each read and its designed variant using several methods based on the barcode. The matching is either edit-distance-based, or uses a faster approximation approach. Then, according to this matching, the tool collects data on the library and the errors that occurred either in the synthesis or the sequencing process. The tool analyzes these statistics and provides a customizable report - plots that summarize this evaluation and analysis of the errors observed. SOLQC also outputs a csv file with the analysis results. The statistics reported by SOLQC include, for example, **Composition Statistics represented by a** histogram and a csv report of the number of reads per variant (Fig. 1) and a sorted bar plot of the number of reads per variant. Another example is the error rates for deletions, substitutions and insertions.

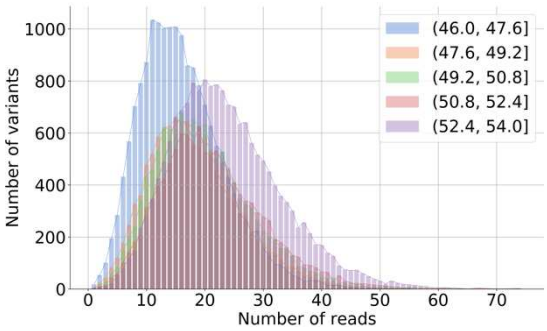


Figure 1: Symbol dependent error rates in [1].

Our presentation will demonstrate SOLQC results for several synthetic DNA libraries that were synthesized and sequenced by different technologies. We compare between these libraries (e.g. Fig. 2) and evaluate the effect of different filtering levels of the reads.

SOLQC is available through a Github download [4] and can be used to analyze synthetic libraries, including for the comparison of different conditions.

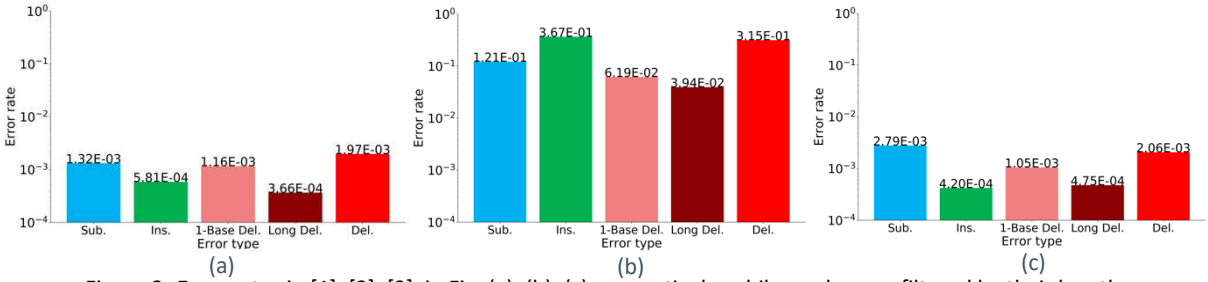


Figure 2 :Error rates in [1], [2], [3], in Fig. (a), (b), (c), respectively, while reads were filtered by their lengths.

[1] Y. Erlich and D. Zielinski. Dna fountain enables a robust and efficient storage architecture. Science, 355(6328):950–954, 2017.
[2] S.M.H.T. Yazdi, R. Gabrys, and O. Milenkovic. Portable and error-free dna-based data storage. *Scientific reports*, 7(1):5011, 2017.
[3] L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, and K. Strauss. Scaling up dna data storage and random access retrieval. bioRxiv, page 114553, 2017.
[4] https://app.gitbook.com/@yoav-orlev/s/solqc/

Super-resolution imaging of single proteins in molecular complexes in cells with small labelling probes and programmable DNA molecules

Thomas Schlichthaerle[1,2], Maximilian T. Strauss[1,2], Florian Schueder[1,2], Alexandra S. Eklund[1,2], Alexander Auer[1,2], Bianca Nijmeijer[3], Moritz Kueblbeck[3], Vilma Jimenez Sabinina[3], Jervis V. Thevathasan[3], Jonas Ries[3], Christian Tiede[4,5], Alistair Curd[4,5], Michelle Peckham[4,5], Darren Tomlinson[4,5], Jan Ellenberg[3] & Ralf Jungmann[1,2]

[1] Faculty of Physics and Center for Nanoscience, Ludwig Maximilian University, Munich, Germany. [2] Max Planck Institute of Biochemistry, Martinsried, Germany. [3] Cell Biology and Biophysics Unit, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany. [4]Astbury Centre for Structural and Molecular Biology, University of Leeds, Leeds, United Kingdom. [5]School of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom.

Optical super-resolution techniques allow fluorescence imaging below the classical diffraction limit of light. From a technology standpoint of view, recent methods are approaching molecular-scale spatial resolution. However, this remarkable achievement is not easily translated to imaging cellular components, as current labelling approaches are limited by either large label sizes (antibodies) or the sparse availability of small and efficient binders (nanobodies, aptamers, genetically-encoded tags). Additionally, it still remains a challenge to extract precise molecule numbers from the analysis of fluorophore photo physics. DNA-PAINT (points accumulation for imaging in nanoscale topography) uses the transient hybridization of programmable fluorophore-labelled oligonucleotides to their respective target sequence to create the blinking pattern typically observed in single-molecule localization microscopy techniques (SMLM). Although 5 nm resolution was previously shown on DNA Origami structures, it's true potential in cellular imaging can only be unravelled with small labelling probes. Here we show the use of small new protein binders, called Affimers, for super-resolution actin imaging in 3D and the application of genetically encoded tags (Halo & SNAP-Tag) to unravel the role of single proteins in the nuclear pore complex with a precision of ~3 nm. We show, that Halo tagged NUP96-proteins are arranged in a cytoplasmic and a nuclear ring, which are 61 nm apart as well as measure the distance between two single NUP96 proteins arranged in the Y-complex in one of the 8 symmetry centres on one ring to be 12 nm apart. The technological advancement of using programmable DNA molecules for super-resolution microscopy now enables structural studies of multicomponent complexes on the level of single proteins in cells.

pH responsive DNA origami lipid nanopore

Jasleen Kaur Daljit Singh, Natalie Surace, Minh Tri Luu, Joel Hochstetter, Eric Jenner, Charlotte Haunton, Claudia Leung, Ravi Jaiswar, Michael Liu, Zhenxu Yang, Ali Abbas, Shelley Wickham (✉ shelley.wickham@sydney.edu.au)

Despite significant enhancements in drug delivery systems, efficient trafficking of drugs into cells remains a challenge in therapeutics. The majority of biological agents, such as peptides, proteins and siRNAs are uptaken into cells via the endocytic pathway. Generally, these molecules are entrapped within the endosomes, where the pH decreases from physiological pH of 7.4 to a pH of 5.5. Endosomal entrapment results in the recycling of the molecules back to the cell surface or in their enzymatic degradation in the lysosome. This presents a significant difficulty in delivering therapeutic agents to the cytosol. Thus, a rate limiting step in the efficient delivery of therapeutic agents is their endosomal escape.

Due to the advances in structural DNA nanotechnology, a large variety of lipid-interacting DNA nanostructures including synthetic lipid membrane channels [1] have been developed in the past decade. As such, to overcome the problem of endosomal escape, we have designed a system for a pH responsive DNA lipid nanopore. We have taken a 6-helix bundle DNA nanopore [2] and enclosed it within a protective DNA origami barrel. This DNA nanopore is decorated with cholesterol moieties to allow it to embed into the lipid membrane, creating a pore. The nanopore is held within the DNA barrel by a scaffold tether. In its 'switched-off' state the scaffold tether is shortened by 'cinch' staple strands, holding the nanopore tightly inside the barrel. The tether can be lengthened by strand displacement of the cinch staples. We have also designed a pH responsive tether incorporating an i-motif, which will be triggered by the acidic conditions of the endosome to lengthen the tether. Once the tether is lengthened, the nanopore is 'switched on' and can find its way out of the barrel, where it is designed to be able to insert into a lipid membrane.

References

[1] M. Langecker et al., "Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures," Science, vol. 338, no. 6109, pp. 932–936, Nov. 2012.

[2] J. R. Burns, E. Stulz, and S. Howorka, "Self-Assembled DNA Nanopores That Span Lipid Bilayers," Nano Lett., vol. 13, no. 6, pp. 2351–2356, Jun. 2013.

Reversible Models for Programming Molecular Computers

William Earley
w.earley@damtp.cam.ac.uk
Department of Applied Mathematics and Theoretical Physics
University of Cambridge

A number of approaches to building molecular computers exist, including the Tile-Assembly Model (TAM [1]) and DNA Strand Displacement (DSD [2]). In addition, robust programming models and tools have been developed, such as Visual DSD [3]. Here, we consider what computational models may be appropriate for a hypothetical class of reversible molecular computers based on term-rewriting systems. Molecular computers, particularly those based on nucleic acids, are good substrates for term-based systems due to their ability to represent (nearly) any sequence, combined with their ability to bind in a sequence-specific manner. Rothemund [4] proposed an early design of a computer based on this observation, specifically a Turing-Machine, taking inspiration from the resemblance of a length of DNA to a Turing tape. Thus we believe that a term-based model incorporating both the sequence nature and association capabilities (sequence-specific recognition) is a useful abstraction. Furthermore, we focus on reversible instances of these models (though adding irreversible primitives is feasible if desired). Given that the laws of physics are fundamentally time reversible and that molecules typically exist in a thermal environment, it is difficult to isolate these systems from noise, and so construction of irreversible molecular computers is challenging. Therefore, fundamentally reversible DNA computers should be a useful avenue to explore.

We present a computational model for describing and programming such reversible systems, the ℵ calculus, and accompanying language `alethe` (roughly meaning *not forgotten* in Greek). ℵ is inspired by the λ calculus and declarative languages such as `prolog`. It is a term rewriting system whose terms are trees and whose leaves are atoms. We then define partial bijections between sets of terms, and these bijections describe processes of 'knowledge transformation' in which knowledge of input variables is reversibly forgotten in exchange for learning the values of the output variables. These partial bijection rules must satisfy certain constraints in order to ensure reversibility in the microscopic sense. The rules may also map between multiple terms, and thus natively support concurrency. The ℵ calculus is proved to be Turing complete by implementing the Reversible Turing Machines described by Bennett [5], and also by implementing the μ-recursive functions. Finally, we also give a semantics for ℵ, and develop a derived programming language `alethe` with an accompanying interpreter and a library of example code.

References

[1] Erik Winfree. *Simulations of Computing by Self-Assembly*. Tech. rep. California Institute of Technology, Department of Computer Science, 1998.

[2] Georg Seelig et al. 'Enzyme-free nucleic acid logic circuits'. en. In: *Science* 314.5805 (Dec. 2006), pp. 1585–1588. ISSN: 0036-8075, 1095-9203.

[3] Matthew R Lakin et al. 'Visual DSD: a design and analysis tool for DNA strand displacement systems'. In: *Bioinformatics* 27.22 (2011), pp. 3211–3213.

[4] Paul Wilhelm Karl Rothemund. 'A DNA and restriction enzyme implementation of Turing Machines'. In: *DNA based computers* 27 (1995), pp. 75–119.

[5] Charles H Bennett. 'Logical Reversibility of Computation'. In: *IBM J. Res. Dev.* 17.6 (Nov. 1973), pp. 525–532.

Performance Trade-offs in Reversible Amorphous Computers

William Earley
w.earley@damtp.cam.ac.uk
Department of Applied Mathematics and Theoretical Physics
University of Cambridge

An important question when designing computers is how performant they can be. Whilst much of this comes down to good engineering, it is also important to determine what limits the laws of physics place on computational performance. Two important such results reveal a quantum mechanical processing limit of $2c^2/h = 2.71 \times 10^{50}\,\mathrm{bit\,s^{-1}\,kg^{-1}}$ [1], and a thermodynamic constraint that each erased bit has a minimum energy cost of $k_b T \log 2$ [2]. As almost all models of computation are logically irreversible, this minimum energy cost is very relevant and a simple geometric argument shows that the size of these systems must asymptotically scale as their (convex) surface area, A.

Reversible computing [3, 4] provides an alternative approach that theoretically allows computation without dissipation. In practice, thermal coupling prevents this, but it is still possible to obtain significantly improved scaling in comparison to irreversible computers. Molecular computers are perhaps uniquely placed to fit this niche, and so we develop an understanding of the performance constraints that affect reversible molecular and amorphous computers from three different perspectives.

Firstly, we prove an empirical observation of Frank [5] about the optimal entropy cost per computational step, and use this to demonstrate that these computers can usefully scale with their volume, rather than area. It is found that the net number of computational state transitions per unit time, R, can scale as $R \sim \sqrt{AV} \sim V^{5/6}$, where V is the system's volume. Numerical estimates show that even very conservative biological computers may significantly outperform contemporary computers at practically all length scales by this metric. In order to achieve this, however, each constituent computational entity must run asymptotically slowly, with a forward bias b that scales as $b \sim V^{-1/6}$, and thus each individual computational entity may well be very slow.

Secondly, we consider the consequences of interacting and communicative processes within this low-bias thermal regime. Using a Fokker-Planck approach, we find that each interaction between two distinct entities results in a time penalty on the order of $1/b^2$, compared to the time for an isolated entity to make one step, $1/b$. This is a significant obstacle, and will be an important consideration for large scale reversible computers.

Thirdly, we evaluate strategies for distributing shared resources such as memory between computational entities. Passive approaches comparable to the biological provision of monomers such as amino acids and nucleotides are found to be impractical or even injurious in the low bias regime. In contrast, active approaches—in which the individual computational entities directly manipulate the resources themselves—are found to be fully compatible with such systems.

References

[1] Norman Margolus and Lev B Levitin. 'The maximum speed of dynamical evolution'. In: *Physica D* 120.1 (Sept. 1998), pp. 188–195.

[2] Rolf Landauer. 'Irreversibility and heat generation in the computing process'. In: *IBM J. Res. Dev.* 5.3 (1961), pp. 183–191.

[3] Charles H Bennett. 'Logical Reversibility of Computation'. In: *IBM J. Res. Dev.* 17.6 (Nov. 1973), pp. 525–532.

[4] Edward Fredkin and Tommaso Toffoli. 'Conservative Logic'. en. In: *Collision-Based Computing*. Ed. by Andrew Adamatzky. Springer London, 1981, pp. 47–81.

[5] Michael Patrick Frank. 'Reversibility for efficient computing'. PhD thesis. Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, 1999.

Cumulative deformation of a linear DNA origami nanoarm designed by module-based approaches

Kohei Mizuno[1], Yuki Suzuki[2,*], Ibuki Kawamata[1], Satoshi Murata[1]
[1]Department of Robotics, Graduate School of Engineering, Tohoku University, Japan
[2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Japan
{mizuno, ysuzuki79r, kawamata, murata}@molbot.mech.tohoku.ac.jp

Soft and flexible movements seen in biological systems are often produced through the cumulative action of the repetitively aligned basic mechanical units as represented by motion segments of spines and sarcomeres of muscles. Mimicking such a cumulative strategy at the single nanostructure level would be a promising route to devising nanomachines that exhibit a smooth, lifelike, motion; however, this attempt intrinsically requires miniaturization of the mechanical units, which could be achievable only by self-assembly from molecules. Here, we designed a linear DNA origami nanoarm that consists of repeats of a tension-adjustable module (Fig. 1 A). Each module can be bent by additional DNA strand having a single-stranded linker region flanked with staple sequences (Fig. 1 B). The linkers act as entropic springs that applied a tension inducing the module to bend. The cumulative actuation of these modules results in a large deformation of the nanoarm, which shift from linear shape into an arched shape. We demonstrate that the degree of deformation is systematically controlled by changing the length of the linker (Fig. 1 C). Moreover, by employing the G-quadruplex-forming sequences for the actuation, we realized K^+-responsive reversible deformation of the origami structure (Fig. 1 D). Our module-based method provides a versatile approach in the design and construction of DNA origami nanostructure that is deformable in response to external stimuli.

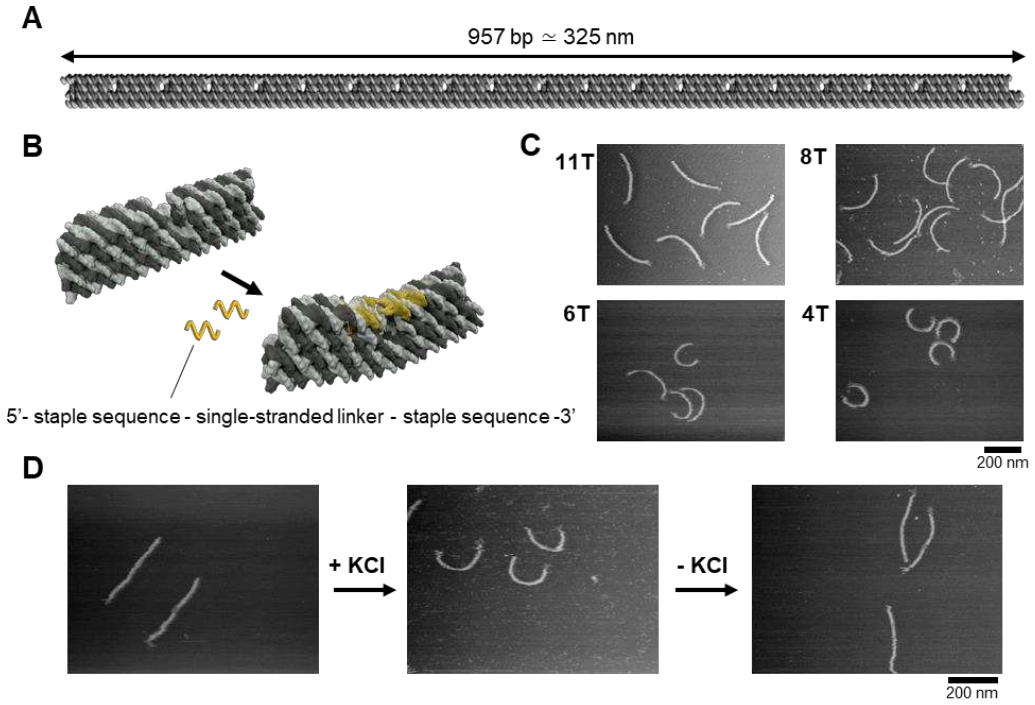


Figure 1. (A) Schematic drawing of the linear DNA origami nanoarm. (B) Schematic drawing of the tension-adjustable module. (C) AFM images of different versions with 11T, 8T, 6T, and 4T linkers. (D) AFM images of the K^+-responsive reversible deformation.

References

[1] P. W. Rothemund, *Nature* **2006**, *440*, 297-302.
[2] S. M. Douglas, H. Dietz, T. Liedl, B. Hogberg, F. Graf, W. M. Shih, *Nature* **2009**, *459*, 414-418.
[3] H. Dietz, S. M. Douglas, W. M. Shih, *Science,* **2009**, *325*, 725-730.
[4] L. Zhou, A. E. Marras, H. J. Su, C. E. Castro, *ACS Nano*, **2014**, *8*, 27-34.

DNA nanostructures from double-C-shaped motifs with controllable angles, twist and curvature

Kai Huang[1, †], Donglei Yang[2, 3, †], Zhenyu Tan[1], Silian Chen[4, 5], Pengfei Wang[3], Ye Xiang[4], Yongli Mi[6], Chengde Mao[7, *], Bryan Wei[1, *]

[1]*School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China*

[2]*School of Chemical Science and Engineering, Tongji University, Shanghai 200092, China*

[3] *Institute of Molecular Medicine (IMM), Renji Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China*

[4]*Center for Global Health and Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Beijing Advanced Innovation Center for Structural Biology, Department of Basic Medical Sciences, School of Medicine, Tsinghua University, Beijing 100084, China*

[5]*School of Life Sciences, Peking University, Beijing 100084, China*

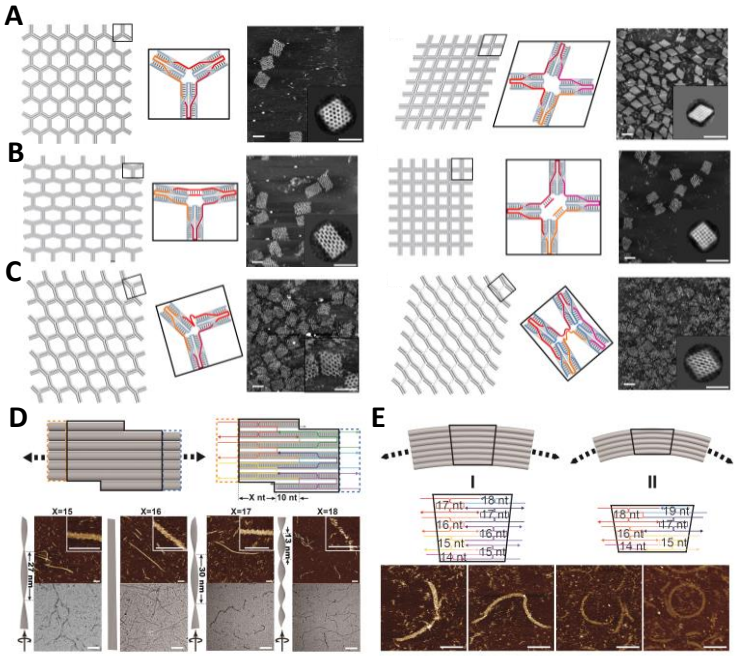
[6] *Department of Chemical and Biological Engineering, the Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR*

[7]*Departmetn of Chemistry, Purdue University, West Lafayette, Indiana 47907, USA*

[†]*These authors contributed equally. *Correspondence and requests for materials should be addressed to B.W. (email: bw@tsinghua.edu.cn) or to C.M (mao@purdue.edu)*

Abstract

As a latecomer in DNA nanotechnology, DNA origami approach has enabled all kinds of DNA nanostructures, and quite a few sophisticated design schemes have been carefully investigated. Similar design frameworks such as wireframe architecture, twist and curvature control have yet to be fully explored in scaffold-free 'LEGO' approach. In this study, we have demonstrated a general design framework to construct DNA nanostructures with controllable angles, twist and curvature from double-C-shaped motifs. Firstly, we design X-shaped and Y-shaped motifs, derivatives of the original double-C-shaped motifs, to form 2D addressable wireframe nanostructures. Basing on the 2D wireframe structures, we demonstrate that single-stranded or double-stranded linkers at certain vertex can result in a desired angle control between the corresponding arms. Twist and curvature can also be engineered by targeted deletion or insertion of base pairs at chosen segments of the double-C-shaped motifs. In doing so, we construct extended ribbons with different levels of twist and curvature.



Summary figure. DNA nanostructures based on double-C-shaped motifs. (**A**) 2D addressable wireframe structures from X-shaped and Y-shaped armed motifs. (**B**) 2D addressable wireframe structures after incorporation of double-stranded linkers. Left panel: from Y-shaped vertices (Figure A left panel) to T-shaped vertices; right panel: from X-shaped vertices (Figure A right panel) to cross-shaped vertices. (**C**) 2D addressable wireframe structures after incorporation of single-stranded linkers. Left panel: from Y-shaped vertices (Figure A left panel) to irregular 3-arm vertices; right panel: from X-shaped vertices (Figure A right panel) to H-shaped vertices. (**D**) Extended ribbons with different levels of global twist. (**E**) Extended ribbons with different levels of global curvature. Scale bars: 100 nm.

Possibility and impossibility in DNA strand displacement

Robert F. Johnson and Lulu Qian

The Chemical Reaction Network (CRN) model describes a system of abstract chemical species and their reactions, and is a common candidate programming language for molecular systems. DNA strand displacement (DSD) is a set of simple yet powerful mechanisms using short DNA strands, and has been used to build a number of devices including various schemes for implementing arbitrary CRNs. Recently, work on formalizing DSD mechanisms has allowed algorithmic enumeration of DSD systems. We suspect that this can further lead to formal analysis of and proofs about the capabilities of DSD systems, and we give a set of restrictions under which arbitrary CRNs cannot be implemented [1]. We further suspect that this formal analysis will help with systematic design of DSD systems, and we present a simplified DSD implementation of arbitrary CRNs that was discovered in the process of that proof [2]. For further references, see [1].

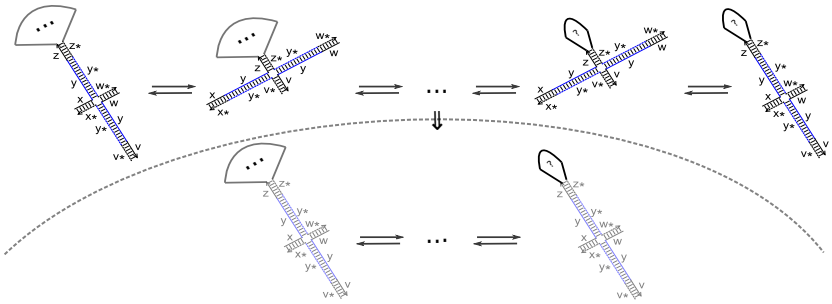


Figure 1: We showed in [1] that with the given restricted set of mechanisms, for any sequence of unimolecular reactions that breaks and reforms a 4-way junction, there is a sequence with the same result that never breaks the junction. This locality theorem is central to the impossibility result.

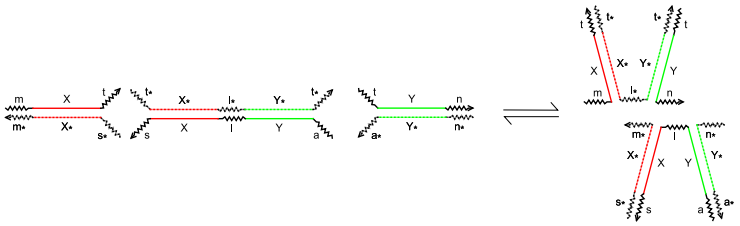


Figure 2: Allowing trimolecular reactions but meeting all other restrictions from [1], this speculative cooperative 4-way branch migration reaction can be used to implement arbitrary reversible CRNs [2].

References

[1] Robert F Johnson. Impossibility of sufficiently simple chemical reaction network implementations in DNA strand displacement. In *International Conference on Unconventional Computation and Natural Computation*, pages 136–149. Springer, 2019.

[2] Robert F Johnson and Lulu Qian. Simplifying chemical reaction network implementations with two-stranded DNA building blocks. In preparation.

Designing switchable Cas12a guide RNAs

Lukas Oesinghaus[1], Friedrich C. Simmel[1]

[1]*Physics Department E14, Technical University Munich, 85748, Garching*
Email: lukas.oesinghaus@ph.tum.de

Nucleic acid strand displacement is one of the most widely used processes in molecular computing owing to its well-understood nature and designability [1]. However, applications are limited by the availability of suitable interfaces to other functional systems. CRISPR-associated proteins are controlled by nucleic acid inputs and have proven to be powerful tools in genome editing, transcriptional regulation, and molecular diagnostics. Cas12a, specifically, has two features that differentiate it from the more commonly used Cas9: (i) It processes its own guide RNAs (gRNAs) and (ii) has an unspecific ssDNase activity that can be used for DNA detection [2, 3].
Here, we control the activity of Cas12a by designing extended gRNAs that are switchable via nucleic acid strand displacement, called strand displacement gRNAs (SD gRNAs). By occluding the handle of the gRNA with a switch domain, we suppress binding of Cas12a (**Fig. 1A**). A single-stranded RNA trigger can displace this switch domain via a toehold, restoring the handle structure and promoting Cas12a binding. Cas12a then processes the SD gRNA, restoring a regular-length gRNA with full activity.
Using NUPACK's constrained multistate sequence design feature [4], this concept can be extended to orthogonal activation by different triggers (**Fig. 1B**), multi-input SD gRNAs similar to the work by Green *et al.* [5] (**Fig. 1C**), and sensing of natural RNA sequences such as mRNAs. It can also be adapted to transcriptional repression using DNase-deactivated Cas12a in *E. coli*, where activated SD gRNAs show repression efficiencies equivalent to regular gRNAs (**Fig. 1D**) [6].
Combining strand displacement with the abilities of Cas12a enables a range of new applications in sensing and synthetic biology, but also introduces a number of interesting design challenges that are uncommon for traditional strand displacement systems.

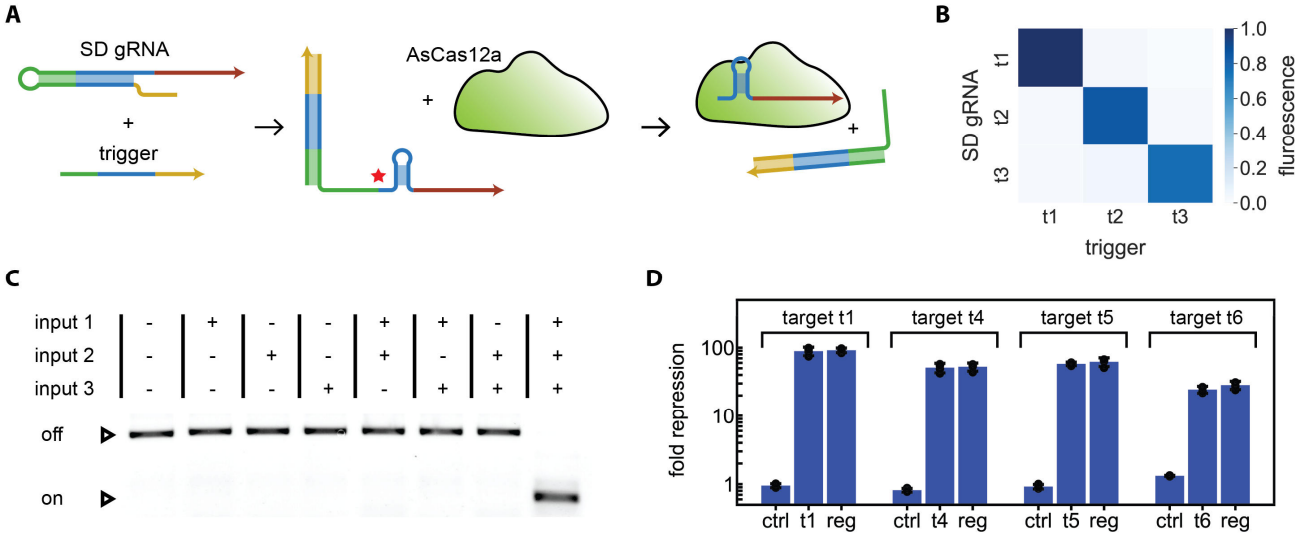


Figure 1. **A** The principle of strand displacement guide RNAs. The red star indicates the position at which Cas12a cleaves the SD gRNA. **B** Orthogonal activation of SD gRNAs by different trigger RNAs as measured by Cas12a's ssDNase activity. **C** A three-input SD gRNA AND gate. **D** Using SD gRNAs for transcriptional repression of mVenus for four different target sequences in *E. coli*. (ctrl: control trigger, t1-t6: cognate trigger, reg: regular gRNA)

References

[1] D. Y. Zhang *et al.*, *Nat Chem* **3**, 103-113 (2011).
[2] I. Fonfara *et al.*, *Nature* **532**, 517–521 (2016).
[3] S. Chen *et al., Science* **360**, 436–439 (2018).
[4] B. R. Wolfe *et al., JACS* **139**, 3134–3144 (2017).
[5] A. A. Green *et al.*, *Nature* **548**, 117–121 (2017).
[6] L. Oesinghaus *et al.*, *Nature communications* **10**, 2092 (2019).

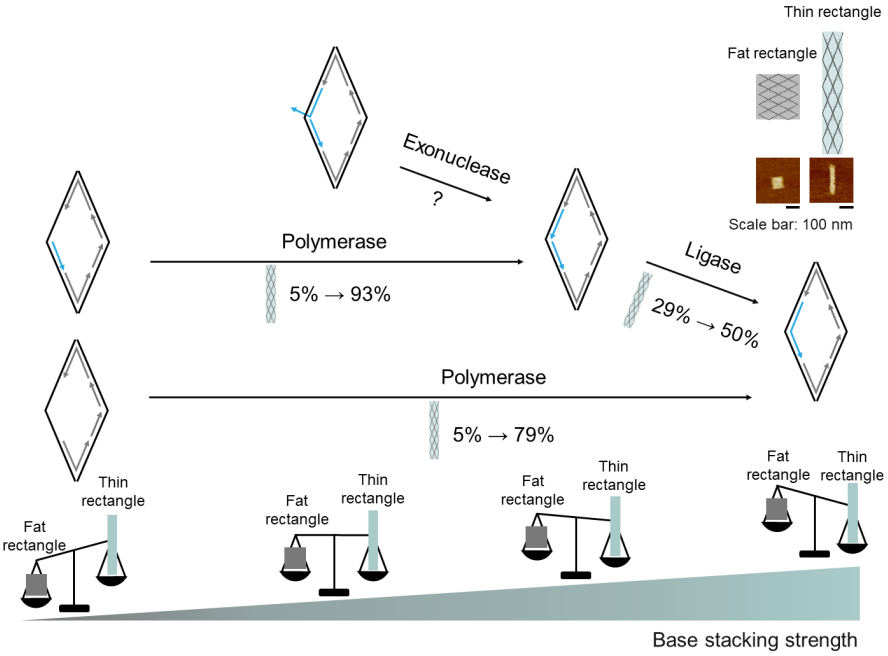
Conformational isomerization of DNA nanostructures based on enzymatic treatment

Qi Yan, Yaqi Wang, Bryan Wei*

School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

*Correspondence and requests for materials should be addressed to B.W. (email: bw@tsinghua.edu.cn)

In a specially designed DNA origami nanostructure, the local conformational change of a few individual junctions can lead to a global conformational isomerization[1,2]. Two conformational isomers with different aspect ratios, the fat rectangle and the thin rectangle, are available in our dynamic DNA nanostructure system. The local conformation at the boundary positions plays a crucial role to induce a certain global conformation. According to our careful investigation, a specific local conformation is determined by the base stacking orientation and strength of tie staples at boundaries, and different types of tie staples have different levels of stacking strength. Continues tie staples, nicked ones and ones with overhangs have decreasingly weaker stacking strength. When enzymes are applied to trim tie staples from one type to another, the strengthened stacking can direct the desired conformational isomerization (e.g. from fat rectangle to thin rectangle). As shown in the summary figure, we present the successful conformational isomerization from a number of enzymatic treatments.



Summary Figure. Enzymatic treatment map based on stacking strength gradients of different types of tie staples. The enzymatic treatment on tie staples (shown in blue) results in the desired conformational isomerization to thin rectangle.

References

1 Song J, Li Z, Wang P, et al. Reconfiguration of DNA molecular arrays driven by information relay[J]. Science, 2017, 357(6349): eaan3377.

2 Cui Y, Chen R, Kai M, et al. Versatile DNA origami nanostructures in simplified and modular designing framework[J]. ACS nano, 2017, 11(8): 8199-8206.

Construction of DNA Amplification Circuit for Directing DNA Nanodevices and Quantifying Nucleic Acids

Ken Komiya, Teruya Enomoto and Masayuki Yamamura
School of Computing, Tokyo Institute of Technology
4259, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa, 226-8503, Japan
komiya@c.titech.ac.jp

Development of a modular reaction that amplifies a pre-designed, single-stranded DNA (ssDNA) as a signal transducer is an essential demand for reliable DNA-based computing and DNA-based robotic systems since it enables signal restoration and sophisticated nanodevice operation following DNA-based computing. In contrast to the conventional ssDNA amplification reaction [1], a short DNA sequence was avoided in our reaction design for allowing stable hybridization of the amplified ssDNA to its target during amplification even under low-temperature conditions. In this context, low temperature is defined as the temperature that falls below the melting temperature (T_m) of the DNA sequence to be amplified. However, a high amplification rate is necessary for supplying a large amount of ssDNA to direct a swarm of nanodevices in response to a tiny amount of nucleic acid stimuli. Although a number of enzymatic and non-enzymatic ssDNA amplification reactions have been developed so far, those achieving efficient amplification and concurrent hybridization of the amplified ssDNA under low-temperature conditions remain rare [2].

In the present study, we constructed an isothermal cascade reaction as a sensor and signal amplifier module that efficiently amplifies ssDNA. Based on the finding that non-specific DNA amplification can be suppressed with the use of locked nucleic acid, we achieved up to million-fold ssDNA amplification and concurrent hybridization to its target at a physiological temperature in one pot and in giant unilamellar vesicles [3, 4]. In addition, the present amplification reaction that exhibits high correlation between molecular concentration and amplification time is applicable for nucleic acid quantification [5].

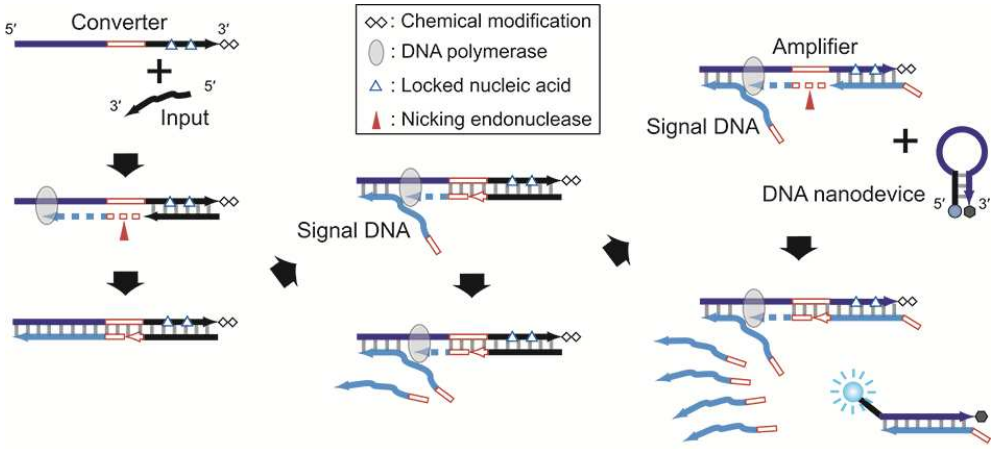


Fig. 1. The scheme of low-temperature amplification (L-TEAM) reaction.

References

[1] J. Van Ness, L. K. Van Ness, D. J. Galas, *Proc. Natl. Acad. Sci. USA*, 100(8):4504-4509 (2003)
[2] X. Chen, N. Briggs, A. D. Ellington, *et al.*, *Proc. Natl. Acad. Sci. USA*, 110(14):5386-5391 (2013)
[3] K. Komiya, S. Kobayashi, M. Yamamura, *et al.*, *Org. Biomol. Chem.*, 17(23):5708-5713 (2019)
[4] Y. Sato, S. M. Nomura, *et al.*, *Chem. Commun.*, *in press* (2019) DOI: 10.1039/c9cc03277k
[5] M. Komori, K. Komiya, T. Yoshimura, *et al.*, *Anal. Bioanal. Chem.*, 411(17):3789-3800 (2019)

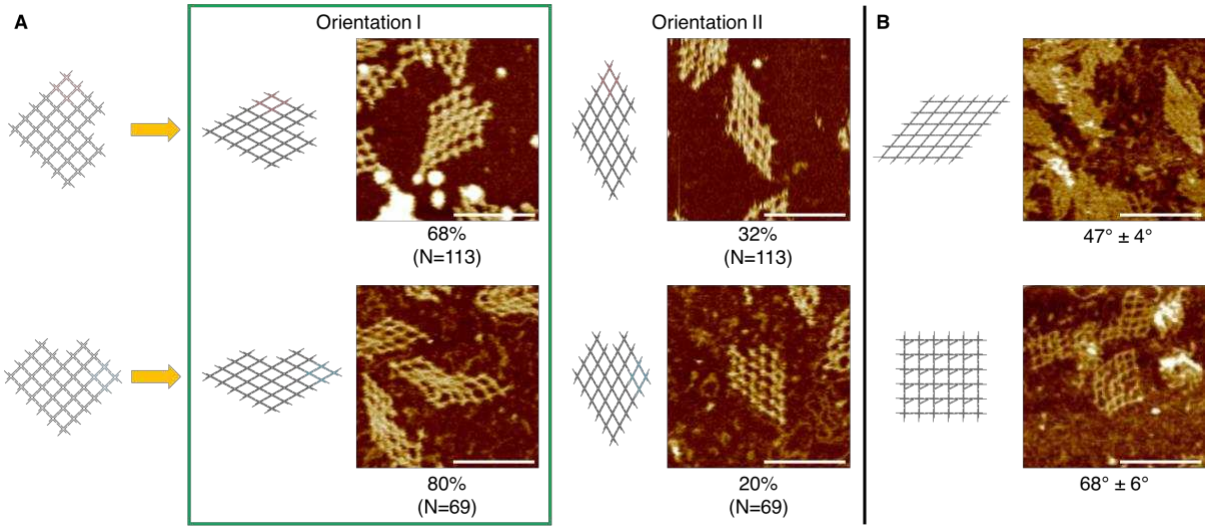
Angle Control of 2D Tessellation Patterns by Simple Junction Motifs

Wen Wang, Tianqing Zhang, Chunyu Chen, Chengxian Wu, Bryan Wei*

School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

*Correspondence and requests for materials should be addressed to B.W.
(email: bw@tsinghua.edu.cn)

Programmed self-assembly of DNA motifs has enables a variety of nanoscale tessellation patterns, which can in turn serve as high-fidelity template to host different guest molecules. Upon revisit of the tessellation patterns from 4-arm junction motifs, we find that the 4-arm motif take an X-shaped configuration with an angle of around 40° instead of a cross-shaped configuration[1]. In our special design of a 6×6 diamond tessellation pattern with a missing corner, an orientation preference of the diamond lattice is presented. The orientation uncertainty indicates that the 4-arm vertices are flexible. When angle-controlling struts are implemented, a tessellation pattern by 4-arm junction motifs with predetermined angles is achieved. The precisely angle control of the tessellation patterns could expand the toolbox of dynamic DNA nanostructures and enable the construction of rigid 3D tessellation architectures.



Summary Figure A. A preferred angle orientation was found of 4-arm tessellation patterns. B. 4-arm tessellation patterns with predetermined angle are formed with the implementation of angle-controlling struts. (Scale bars: 100 nm)

Reference:

1 Sha, R., Liu, F. & Seeman, N. C. Atomic Force Microscopic Measurement of the Interdomain Angle in Symmetric Holliday Junctions. *Biochemistry* **41**, 5950-5955(2002).

Automated Design of Scaffold-free Wireframe DNA Nanostructures

Abdulmelik Mohammed[1], Wei Wang[2], Bryan Wei[2] and Pekka Orponen[1]

[1] *Department of Computer Science, Aalto University, Finland*
[2] *School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, China*
Emails: abdulmelik.mohammed@aalto.fi, wenwang730@gmail.com, bw@tsinghua.edu.cn, pekka.orponen@aalto.fi

DNA self-assembly, a process whereby a mixture of designed DNA strands self-assemble into custom nanoscale constructs, has emerged as a promising route for bottom-up synthesis of artificial nanostructures with potential applications in areas such as nanomedicine, biophysics and plasmonics. Recently, wireframe architectures, which offer a materially efficient scheme for designing physiologically stable DNA nanostructures, have gained significant attention, especially due to new automated approaches [2, 3] for folding long, circular DNA origami scaffold strands into intricate polygonal mesh wireframes. However, scaffolded origami designs are inherently limited by the lengths of available scaffold strands and the topological complexity of routing such circular strands into more complicated target geometries.

Taking the limitations of scaffolded approaches into account, Wen et al. [1] introduced a general scaffold-free method for designing DNA wireframe nanostructures from short synthetic strands. In this work, we automate the design scheme of Wen et al. [1] for meshes with an underlying space of an *orientable* topological surface, such as the pentagonal torus in Figure 1. Such polygon meshes permit an antiparallel orientation of the polygons' bounding cycles, which in turn allows for single-duplex renderings of the edges from the antiparallel double tracings performed by the cycles. To convert the initial cyclic strands to shorter linear ones, the cyclic strands are nicked in every edge in a sticky-ended pattern. We implemented the method using vhelix [2] as a backend, thus paving a route towards an integrated environment for designing scaffolded and scaffold-free 2D and 3D wireframe DNA nanostructures.

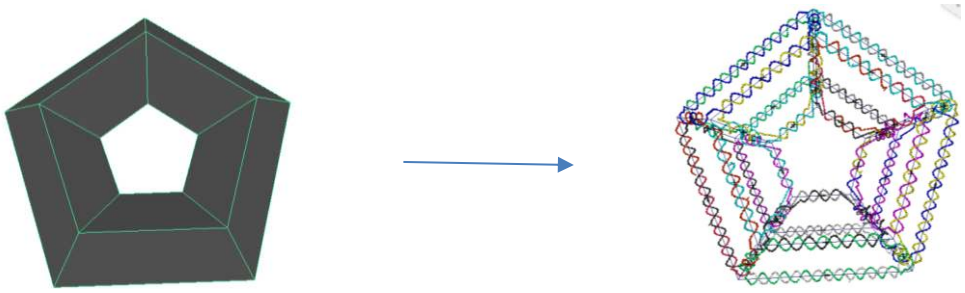


Figure 1: A pentagonal torus 3D model (left) rendered to a scaffold-free wireframe DNA (right).

References

[1] W. Wang, S. Chen, B. An, K. Huang, T. Bai, M. Xu, G. Bellot, Y. Ke, Y. Xiang and B. Wei, Nature Communications. 2019, **10**, 1067.
[2] E. Benson, A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen and Björn Högberg. Nature 2015, **523**(7561), 441—444.
[3] R. Veneziano, S. Ratanalert, K. Zhang, F. Zhang, H. Yan, W. Chiu and M. Bathe. Science 2016, **352**(6293), 1534.

DNA origami nanostructures with scaffolds obtained from rolling circle amplification

Mengyuan Xu[1, †], Chi Zhang[1, †], Cunjun Zhang[1], Chunhai Fan[2], Jie Chao[3, *] and Bryan Wei[1, *]
[1]School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China
[2]School of Chemistry and Chemical Engineering, and Institute of Molecular Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200240, China
[3]Key Laboratory for Organic Electronics and Information Displays (KLOEID), Institute of Advanced Materials (IAM), Jiangsu Key Laboratory for Biosensors, School of Materials Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[†]These authors contributed equally.
[*]Corresponding author. Email: bw@tsinghua.edu.cn (B.W.); iamjchao@njupt.edu.cn (J.C.)

As a guiding backbone of origami structural folding, the long single-stranded scaffold is the key enabler of all kinds of complex DNA nanostructures. However, the folding of most origami nanostructures depends heavily on very limited species of scaffolds. It is desirable to have a wider choice of scaffolds for the extended design space. In this study, rolling circle amplification (RCA) is utilized to produce scaffolds for DNA origami nanostructures based on different templates (Figure 1). 1) RCA product with M13 genomic DNA template directly serves as scaffold to fold origami concatemers. Alternatively, origami monomers can be obtained by restriction enzyme treatment. 2) RCA product with plasmid DNA template is prepared as scaffold to fold into origami concatemers or monomers in a similar way. In general, we present a new strategy based on RCA to prepare custom ssDNA scaffolds that could further empower DNA origami method to produce nanostructures with extended design space. Notably, the 16 kb ssDNA scaffold from a plasmid template is already ~100% larger than a full-size M13 scaffold for typical DNA origami structures.

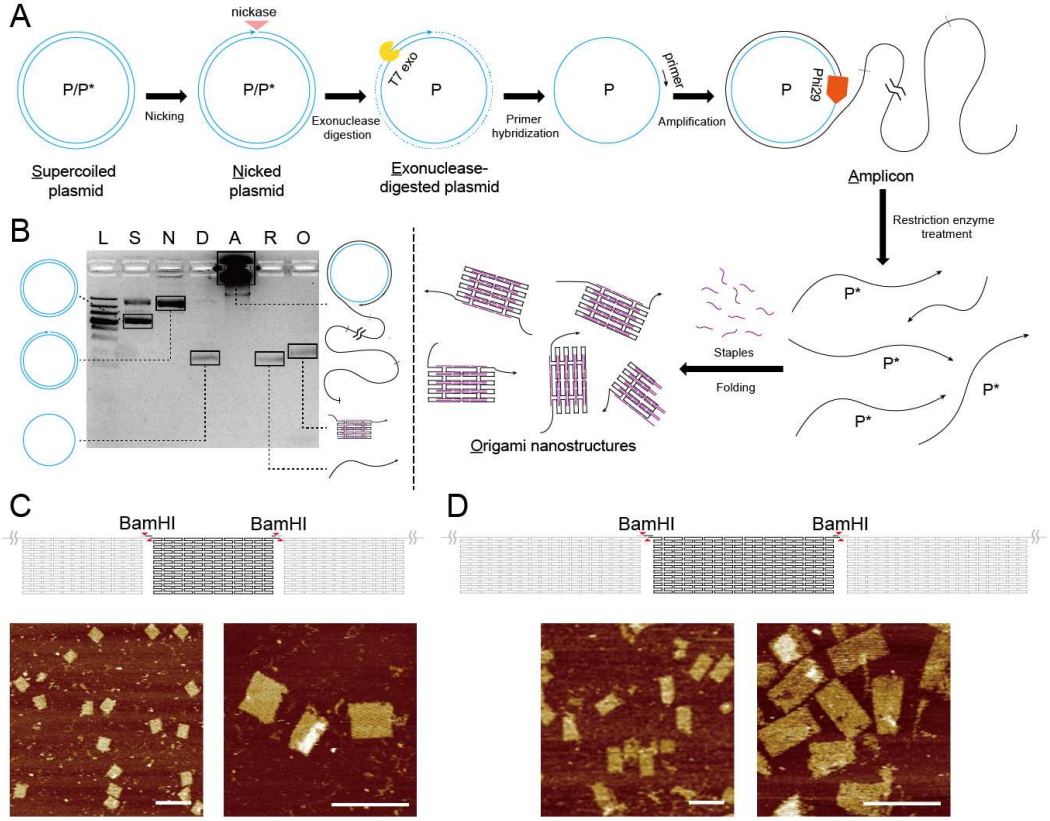


Figure 1. Origami nanostructures folded by scaffold from RCA product based on a plasmid template. (A) Schematic diagram of RCA product with plasmid as scaffold to be used to fold origami structure. (B) Native agarose gel electrophoresis results of the process of different stages. Lane S: circular plasmid dsDNA; lane N: plasmid treated with nicking enzyme; lane E: nicked plasmid treated with exonuclease; lane A: RCA product; lane R: RCA product after restriction enzyme treatment and purification; lane O: folded origami with restriction enzyme treated RCA product as scaffold; lane L; 1 kb DNA ladder. (C and D) Folded origami rectangles with restriction enzyme treated RCA products based on plasmid templates (8 kb plasmid in C and 16 kb plasmid in D). Top: diagrams; bottom: AFM images. Scale bars: 200 nm.

Programmable nucleation of 1D DNA nanostructures with DNA slats

Dionis Minev[1,2,3,4], Chris Wintersinger[1,2,3,4], Richard Guerra[2,3,4], Anastasia Ershova[2,3,4], and William M. Shih[2,3,4]

[1]John A. Paulson School of Engineering and Applied Sciences, Harvard University; [2]Wyss Institute for Biologically Inspired Engineering at Harvard University; [3]Department of Cancer Biology, Dana-Farber Cancer Institute; [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

dminev@g.harvard.edu, william.shih@wyss.harvard.edu

The field of DNA nanotechnology has emerged with a number of powerful self-assembly methods that have proven as extraordinarily useful tools for bottom-up nano-construction. Amongst several, DNA-origami and DNA bricks are the most prominent. In DNA origami, the long single-stranded "scaffold" DNA is mixed with a large excess of short synthetic "staple" oligonucleotides and annealed to create virtually any 2D and 3D nanoscale object. The scaffold interacts with a large number of staples and is used as the limiting component in the assembly. Thus the scaffold controls the nucleation of the assembly, with precisely as many origami structures yielded as there are scaffold molecules. However, the size of each origami is limited by the length of the scaffold (~10kb). DNA bricks, in contrast, utilize short brick-oligonucleotides that are complementary to each other and can assemble into much larger structures independent of a scaffold. This feature conversely limits control over nucleation.

Rapid and nucleation-limited growth is a highly advantageous feature for programmable self-assembly. We present a DNA self-assembly method that allows programmable nucleation of 1D DNA nanostructures via an origami-based "seed" triggering the assembly of oligonucleotides complementary to each other, called DNA "slats". In DNA-slat-based assembly, an origami-seed is necessary to allow the formation of the slats into the desired nanostructure. Exclusion of the seed structure in the reaction results in no assembly. We show that we can trigger seed-dependent assembly of 1D structures under selected salt concentrations, assembly temperatures, and DNA slat concentrations, while suppressing unwanted seed-independent assembly. Furthermore, we present a number of different DNA sequence variants that modulate the monodispersity of the assembled 1D structures. We envision DNA slats as a useful addition to current DNA self-assembly methods, allowing control over the nucleation of nanostructures, while retaining high yields and monodispersity of large 1D structures.

Grant acknowledgments: Wyss Core Faculty Award, Wyss Molecular Robotics Initiative Award, NSF Award 1435964, NSF Award CCF-1317291, ONR Award N000141510073, ONR Award N00014-18-1-2566, BMGF/Ragon Global Health Innovation Partnership Award OPP1126222

Controlled Nucleation of DNA Origami Megastructures using Crisscross Cooperativity

Chris Wintersinger[1,2,3], Dionis Minev[1,2,3], Jonathan F. Berengut[5], William M. Shih[1,3,4]
[1]Wyss Institute for Biologically Inspired Engineering at Harvard University, [2]Harvard John A. Paulson School of Engineering and Applied Sciences, [3]Department of Cancer Biology, Dana-Farber Cancer Institute, [4]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School. [5]University of New South Wales Sydney

Scaffold-based DNA origami has proven powerful for designing and assembling arbitrary nanoscale objects. The method has enabled applications in material science, smart therapeutics, biophysics, and other areas requiring the precise positioning of molecules at nanometer distances. One key limitation of DNA origami is that it is challenging to combine single DNA origamis into multi-component assemblies without also generating undesirable side-products. Sub-components in such assemblies often bind in unintended ways among themselves to cause spurious nucleation of unfinished, erroneous, and aggregated assemblies. To overcome these challenges, we propose crisscross cooperative assembly for bottom-up fabrication of megastructures from DNA origami "slats." Each slat is a DNA origami six-helix bundle nanorod that features single-stranded binding sites extending along the length of the slat. Individual binding sites are short (7–9 bp) such that pairwise interactions between single slats are greatly weakened to eliminate spurious nucleation. Assembly of a desired megastructure requires slats to be first coordinated on a "seed" DNA origami. Once an initial series of slats are bound to the seed, binding sites of several overhanging slats cooperatively engage additional slats one-by-one to extend the megastructure.

Firstly, we show growth of six-helix bundle slats into flexible bi-layer filaments containing over one thousand discrete slats using a gridiron DNA origami seed. These megastructures were frequently tens of micrometers in length and had molecular masses in excess of five gigadaltons. We found reaction conditions where measurable growth of megastructures occurred only when seeds were added to the reaction. Secondly, we determine the average rate of incorporation of slats into megastructures and estimated surprisingly fast kinetics of 10^5–10^6 $M^{-1}s^{-1}$ per slat addition, indicating that crisscross assembly can be used to build megastructures over reasonable timescales. And thirdly, we show that assembly of tri-layer filaments are straighter and more rigid than the aforementioned bi-layer design, showing that mechanical characteristics of the larger megastructure can be determined by varying how slats are linked to one another. We are further refining this system to lessen/eliminate spontaneous nucleation over a broader range of assembly conditions to create a diversity of finite and continuously growing DNA origami slat megastructures.

Grant acknowledgments: Wyss Core Faculty Award, Wyss Molecular Robotics Initiative Award, NSF Award 1435964, NSF Award CCF-1317291, ONR Award N000141510073, ONR Award N00014-18-1-2566, BMGF/Ragon Global Health Innovation Partnership Award OPP1126222

Probing the Physical Limits of Reliable DNA Data Retrieval

Lee Organick[1]*, Yuan-Jyue Chen[2], Siena Dumas Ang[2], Randolph Lopez[3], Karin Strauss[2], and Luis Ceze[1]

[1]University of Washington, Paul G. Allen School of Computer Science and Engineering, Seattle, 98195
[2]Microsoft Research, Redmond, 98052, USA
[3]University of Washington, Department of Bioengineering, Seattle, 98195
*leeorg@cs.washington.edu

Synthetic DNA has been gaining momentum as a potential storage medium for archival data storage. In this process, digital information is translated into sequences of nucleotides and the resulting synthetic DNA strands are then stored for later retrieval. While retrieval of individual digital files via PCR-based random access has been previously demonstrated, the physical limits and reliability of this process have not been thoroughly explored. Previous work recognized the importance of storage density for DNA to become a practical archival storage but did not explore PCR random access accuracy when accessing extremely small subsets of data from a dense, complex pool. First, this work examines the ability of PCR to recover files from pools of far greater complexity, ranging from over 10^6 to over 10^{10} unique sequences per microliter. We refer to these pools as "simple" and "complex", respectively. If PCR file retrieval fails in complex settings, we would observe a marked difference between sequences recovered in complex versus less complex settings. In addition, we might observe an inability to recover small files in complex conditions. Encouragingly, we observed neither of these symptoms after PCR random access on three files of varying size (ranging from 2,042 sequences in the smallest file to 271,447 sequences in the largest file) in both the simple and complex pools. Second, this work examines the minimum physical redundancy needed to successfully recover desired files. Here, we demonstrate reliable file recovery when as few as an average of 10 copies per sequence are stored. This number holds true over all three files examined in both the simple and complex pool conditions discussed above. This results in a density of approximately 17 exabytes/g, nearly two orders of magnitude greater than prior work has shown. Third, this work examines the pattern of sequences missing over simple and complex pool conditions and all dilution conditions. Here we find sequences are not missing systematically and are therefore missing simply due to stochastic variation, thus confirming a robust and reliable sequence design and recovery method. Fourth, we examine the effect of sequencing depth. If a greater sequencing depth is needed, the cost of the sequencing resources might outweigh the benefit of a denser storage system. However, we did not need additional sequencing depth to successfully recover the files under all conditions. These findings substantiate the robustness of PCR as a random access mechanism in complex settings, and that the number of copies needed for data retrieval does not significantly compromise the ultimate density of the system.

Triangular DNA origami nanostructure having multiple self-assembly modes

Yuki TAKEDA[1], Yuki SUZUKI[2], Satoshi MURATA[1]
[1]Department of Robotics, Graduate School of Engineering, Tohoku University, Japan, [2]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Japan
{takeda, ysuzuki79r, murata}@molbot.mech.tohoku.ac.jp

Abstract

Structures or patterns produced through self-assembly are highly dependent on how components interact with each other. Even in the self-assembly of single-component systems, a variety of patterns can be obtained by controlling relative positions and orientations of the components. Here, we designed a triangular DNA origami block exhibiting multiple self-assembly modes based on its self-complementary shape. Each side of the triangular block has protrusions for the connection. The protrusions introduced on the two of the three sides are upside down in relation to each other, while those on the third side are shaped to be complementary to the other two sides (Fig. 1a). By disabling of an arbitrarily selected side, the self-assembly is directed into different patterns such as linear, hexagonal, or dispersed patterns. Both transmission electron microscopy (TEM) and atomic force microscopy (AFM) imaging demonstrated the formation of the desired patterns. The orientation of each triangular component in the assemblies was confirmed by post-assembly modification of the triangle surface with protein molecules, demonstrating the advantage of the building block with well-defined top-down orientation for precise positioning of guest molecules onto the desired side of the assemblies (Fig 1b). We also anticipate that our design approach can be expanded to advanced self-assembly systems whose assembled patterns are interconverted by switching the mode of the components.

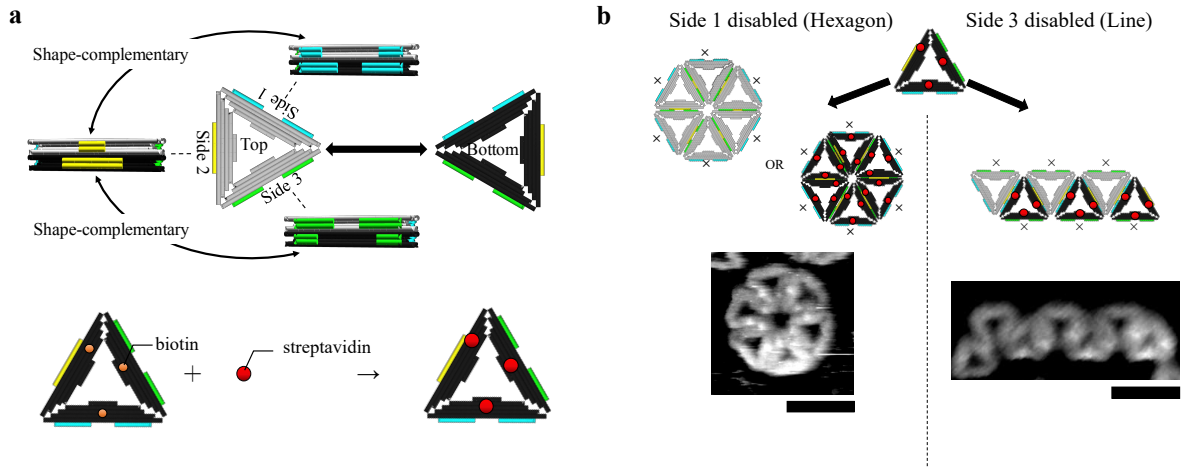


Fig. 1 Schematic diagrams of the triangular DNA origami block. (a) Side 1 and side 3 have protrusions complementary to those patterned on side 2. The pattern of the protrusions on side 1 and that on side 3 are upside down in relation to each other. The connectability of the selected side is disabled by extending staples of the protrusion with poly-T tails which hinder blunt-ended stacking interactions. Biotin molecule was introduced to the bottom (black) surface of the triangle. (b) AFM images of hexagonal and linear shapes after the *in situ* conjunction with streptavidin molecules on mica surfaces. Scale bar: 100 nm.

References

- G. Tikhomirov, *et al., J. Am. Chem. Soc.* **140**, 17361-17354, 2018.
- G. M. Whitesides, *et al., Science* **295**, 2418-2421, 2002.
- P. W. K. Rothemund, *Nature* **440**, 297-302, 2006.
- T. Gerling, *et al., Science* **347**, 1446-1452, 2015.

Dynamics and computation of switchable DNA nanostructures based on toehold-free strand displacement

Hong Kang[1], Tong Lin[2], Xiaojin Xu[2], Qingshan Jia[3], Richard Lakerveld[2], Bryan Wei[1,*]

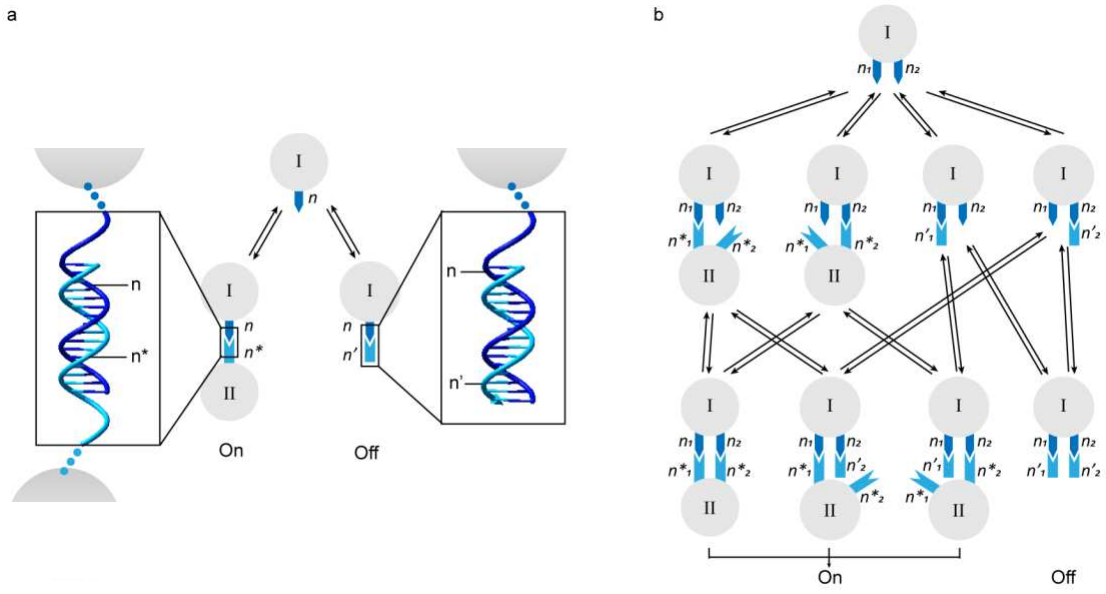
[1]School of Life Sciences, Tsinghua University-Peking University Center for Life Sciences, Center for Synthetic and Systems Biology, Tsinghua University, Beijing 100084, China

[2]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong S.A.R

[3]Department of Automation, Center for Intelligent and Networked Systems, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing, China

*Correspondence and requests for materials should be addressed to B.W. (email: bw@tsinghua.edu.cn)

Instead of using single-stranded overhangs to serve as toeholds to initiate a dynamic structural switch, we have shown in this study that controlled dynamics can be achieved by simpler implementation of toehold-free strand displacement. Under the scheme of toehold-free strand displacement, when excess amount of a blocker strand (e.g. strand n') complementary to either segment of the paired duplex (e.g. n/n*) is introduced, it can block the corresponding segment by pairing competition (e.g. n/n'). The blocking can be applied in a certain set of paired segments (e.g. N/N*, N={n_1, n_2,…, n_i}; N*={n_1^*, n_2^*,…, n_i^*}) respectively. When a set of independent single-stranded DNA blockers (e.g. N', N'={n_1', n_2',…, n_i'}) is added at excess amount, the original pairing scheme will be outcompeted and displaced. In other words, when the original pairing scheme N/N* is defined as an 'on' state, the pairing after competition (N/N') can be defined as an 'off' state. Such an on/off switch is controlled by the addition of blockers at an excess amount. Basic dual-unit systems and a more complex quadruple-unit system are designed based on the concept. The controllable switch between coupling and decoupling in several DNA nanostructure systems are demonstrated. A 2-input and an 8-input Boolean computation are implemented to showcase the controllable dynamics.



Summary Figure. Schematic diagrams of switchable DNA nanostructures based on toehold-free strand displacement. Tree graphs of blocking are shown for coupling/decoupling with one species of blocker (a) or two species of blockers (b).

Binarized Neural Networks as Droplet-Mediated Strand Displacement Cascades

Johannes Linder[1], David Wong[1], Yuan-Jyue Chen[2], Luis Ceze[1], Georg Seelig[1,3], Karin Strauss[2]

[1]*Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA*

[2]*Microsoft Research, Redmond, WA, USA*

[3]*Department of Electrical & Computer Engineering, University of Washington, Seattle, WA, USA*

Neural Nets have recently been demonstrated with DNA strand displacement, using unique DNA species for each weight connection to implement the network as a one-pot cascade [1]. Real-valued weights were implemented by mixing the corresponding weight gates at specific concentrations. Consequently, the number of DNA species and distinct concentrations grow rapidly with network size, posing several challenges for automation and scalability.

Here, we aim to combine spatial separation in microfluidic droplets with a stricter network architecture to address these challenges. By isolating computational primitives in droplets, we can reuse DNA species for all primitives of the same network layer (Figure 1A). We restrict the choice of model to Binarized Neural Nets where all weights are binary-valued (+1 / -1). In this framework, all gates are mixed at identical concentrations. Signals are cascaded through the network by mixing droplets of the current layer with droplets of the next layer (Figure 1B). We ensure orthogonality by switching DNA species at each layer, using 4 unique droplets and 9 DNA species per layer. Simulations indicate that, due to dilution, the same DNA species can be reused after three layers, allowing any network to be built from a constant number of DNA reagents. During training, we optimize neurons to be far away from their decision boundaries, making the network resilient to noise in expectation.

We have experimentally tested a 5-input neuron, using manual pipetting to simulate droplet operations (Figure 1C). While the leak caused problems for patterns close to the decision boundary, we could compute well-separated patterns. Next, we plan on reducing the leak and deploying the network to a microfluidic droplet device (e.g. PurpleDrop).

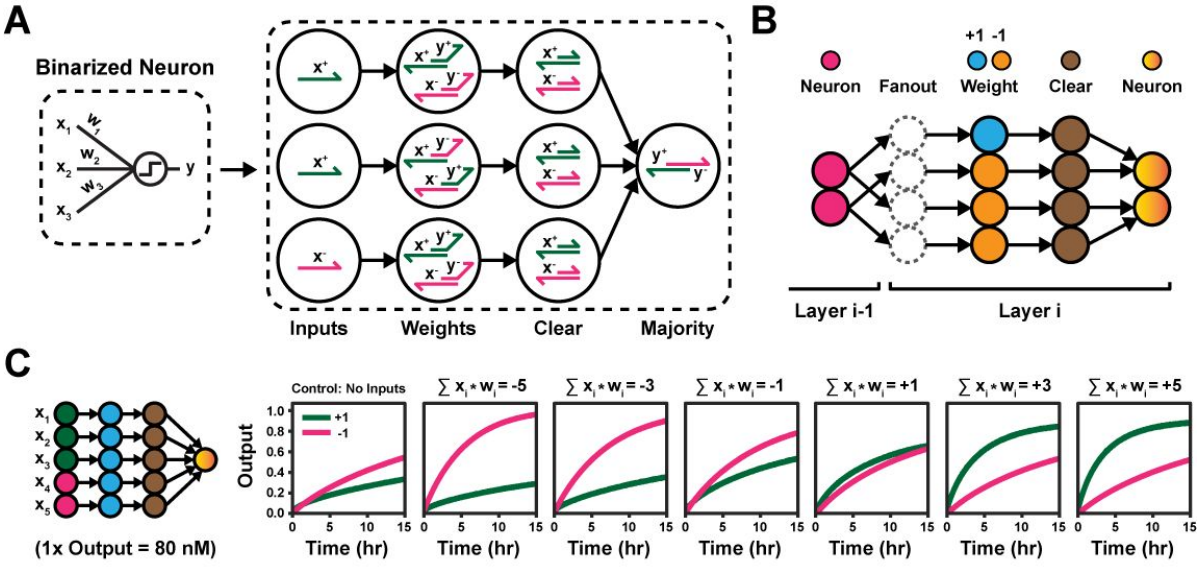


Figure 1. A. The droplet protocol is executed by (1) joining input droplets and catalytic translator weight gate droplets, (2) removing input strands with threshold droplets, (3) merging weighted inputs with an annihilator droplet. **B.** Neuron droplets of the previous layer become inputs to the next layer. **C.** 5-input neuron experiment with catalytic seesaw gates.

[1] Cherry, Kevin M., and Lulu Qian. "Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks." Nature 559.7714 (2018): 370.

ALCH: An Imperative Language for the Chemical Reaction Network-Controlled Tile Assembly Model

Titus H. Klinge[1], James I. Lathrop[2], Sonia Moreno[3], Hugh D. Potter[2], Narun K. Raman[3], and Matthew R. Riley[2]

[1]Drake University, Des Moines, IA 50311 USA, titus.klinge@drake.edu
[2]Iowa State University, Ames, IA 50011 USA, {jil,hdpotter,mrriley}@iastate.edu
[3]Carleton College, Northfield, MN 55057 USA, {morenos,ramann}@carleton.edu

In 2015, Shiefer and Winfree introduced the chemical reaction network-controlled tile self-assembly model (CRN-TAM) to investigate self-assembly systems that interact with non-local chemical signals. Such non-local interaction is observed in nature where organisms commonly communicate through chemical signals. Laboratory experiments also confirm that DNA strand displacement can be used to control structural transformations of DNA origami shapes which is a chemically controlled self-assembly process akin to the model. In this paper, we introduce a programming language called ALCH for specifying CRN-TAM programs and use it to investigate properties of the CRN-TAM, including a decidability result that holds in the individual models but fails in the hybrid model. ALCH is a high-level imperative language that supports standard features such as variables, conditional control, and loops, as well as CRN-TAM specific features and nondeterministic branching. We also developed a software compiler and simulator to translate ALCH programs into reaction-level CRN-TAM programs that can be simulated and visualized in real time.

Using ALCH, we present a strict CRN-TAM construction of the discrete Sierpinski triangle, a well-known fractal that cannot be strictly self-assembled in the aTAM. Our construction builds the lower symmetric half of the DST column by column, building each column tile by tile from the bottom up before starting the next. We face two challenges with this method: we must control specifically where each tile bonds, and we must determine which tile, if any, to add at each position in the column. To ensure that we add each tile only in the correct location, we extend a temporary partial column of scaffold tiles to occlude unintended bond sites. We can then use temporary probe tiles to determine the local structure of the previous column and calculate which tile to add. After we complete each stage of the fractal, we clear out all scaffold tiles from the completed region. We construct the upper symmetric half simultaneously in lock-step, leveraging the probe information from the lower symmetric half.

ALCH and the simulator eased the burden of debugging and helped simplify both our CRN-TAM specification as well as the proof of correctness. We hope that ALCH, the compiler, and the simulator will help accelerate future research and streamline the development process in the CRN-TAM.

Simulation of Tethered One-Legged Molecular Walkers on Independent 1-D Tracks

David Arredondo[2,3] and Darko Stefanovic[1,2,3]
darredondo@unm.edu
[1]Department of Computer Science, [2]Nanoscience and Microsystems Engineering Program, [3]Center for Biomedical Engineering
University of New Mexico

We study the motion of random walkers with residence time bias between first and subsequent visits to a site, as a model for synthetic molecular walkers based on catalytic DNA known as molecular spiders [3, 2, 1, 5]. Previous studies have discovered exceptionally long superdiffusive transients, which would be relevant in experiments and applications. The mechanism of the transient superdiffusion was explained via the emergence of a boundary between the new and the previously visited sites, and the tendency of the multi-legged spider to cling to this boundary, provided there is a residence time bias between the first and the subsequent visits to a site. Detailed results were obtained for a two-legged spider with hand-over-hand gait in one dimension. The key insight is that the spider alternates between being on the visited/new boundary (and moving forward ballistically) and drifting in the sea of previously visited sites; with each period this sea becomes bigger, hence the eventual breakdown of superdiffusion.

Approaches for improving the transient and perhaps asymptotic behavior of molecular spiders include modified chemistry, different walker body and leg geometry, and the use of walker teams. Walker teams are inspired by nature's molecular motor teams [6]. Rank et al. [4] provided a detailed analysis of teams of two-legged molecular spiders, on parallel one-dimensional tracks, connected by a "leash", i.e., a kinematic constraint that no two spiders can be more than a certain distance apart. They showed that teams of two, three, and four spiders successively outperform a single spider, for a range of leash lengths and chemical kinetics.

Here we ask: can we separate the effects of having a team of walkers from the effects of each walker having multiple legs? Our model system uses single-legged walkers, each on its own one-dimensional track, connected by a leash. Each track is prepared with fresh substrates for $x \geq 0$, and consumed products for $x < 0$; walkers start at $x = 0$. Using both kinetic Monte Carlo simulation and an analytical approach, we recapitulate the method of [4]. Even though a single one-legged walker does not exhibit directional, superdiffusive motion, we find that a team of one-legged walkers on parallel tracks, connected by a flexible tether, does enjoy a superdiffusive transient. Furthermore, the one-legged walker teams exhibit a greater expected number of steps per boundary period and are able to diffuse more quickly through the product sea, leading to longer periods of superdiffusion.

References

[1] T. Antal and P. L. Krapivsky. Molecular spiders with memory. Physical Review E, 76(021121), 2007.
[2] K. Lund, A. J. Manzo, N. Dabby, N. Michelotti, A. Johnson-Buck, J. Nangreave, S. Taylor, R. Pei, M. N. Stojanovic, N. G. Walter, E. Winfree, and H. Yan. Molecular robots guided by presciptive landscapes. Nature, 465(7295), 2010.
[3] R. Pei, S. K. Taylor, D. Stefanovic, S. Rudchenko, T. E. Mitchell, and M. N. Stojanovic. Behavior of polycatalytic assemblies in a substrate-displaying matrix. Journal of the American Chemical Society, 128(39), 2006.
[4] M. Rank, L. Reese, and E. Frey. Cooperative effects enhance the transport properties of molecular spider teams. Physical Review E, 87(3):032706, 2013.
[5] O. Semenov, M. J. Olah, and D. Stefanovic. Mechanism of diffusive transport in molecular spider models. Physical Review E, 83(021117), 2011.
[6] A. Vilfan, E. Frey, F. Schwabl, M. Thormählen, Y. Song, and E. Mandelkow. Dynamics and cooperativity of microtubule decoration by the motor protein kinesin. Journal of Molecular Biology, 312(5):1011 – 1026, 2001.

This material is based upon work supported by the National Science Foundation under grant CCF-1422840.

Reservoir Computing for Genome Sequence Classification

Christopher Neighbor Adán Myers y Gutiérrez

Globally, pathogens increasingly threaten human health and welfare. Due to the distribution of disease and the expense of diagnosis, it is of value to develop portable assays that can rapidly detect infection. DNA molecular logic technology offers promise as a portable detection method due to the versatility and stability of DNA and the potential of *in situ* computation. The additional power of DNA is in the potential to perform *in situ* computation in the assay format. Bioassays implementing DNA computation have taken advantage of this capacity to some extent, but one area of biodetection technology that would particularly benefit is next generation sequencing.

The Seelig group recently implemented a DNA-based classifier for expression profiling[2]. The classifier is a support vector machine and inputs are target sequences chosen *a priori*. It would be advantageous to be able to deal with any input without knowing target sequences ahead of time. To this end, we propose a machine learning model which takes a sequence as input, extracts the features through a dynamic reservoir, and then linearly classifies the reservoir nodes' output. Previous work has been done to demonstrate that DNA components can be used to create a reservoir computing system[1]. In our work we explore modeling this chemical reservoir as a fixed recurrent neural network which has been initialized to have particular dynamic properties[3]. A sufficiently dynamic chemical reaction network, which acts as the reservoir, may not need to be constructed for the task, but rather only the readout system need be adapted to the particular task.

We compared the performance of these networks against a benchmark image classification task, a genome intron-exon classification task, and dengue virus serotype classification. As a control, we also tried these tasks using a trainable long short-term memory (LSTM) recurrent neural network of comparable size. The classifier performed well on these tasks achieving accuracy of > 0.95 on each of our tasks which is only a few points away from the fully trained control models.

Dataset	**Input Length**	**Classes**	**Samples**	**Reservoir Acc.**	**LSTM Acc.**
MNIST	28	10	60,000	0.953	0.991
UCI Splice	60	3	3190	0.962	0.972
Dengue	100	4	3625	0.965	0.988

Our eventual goal is to differentiate serotypes, or other classes of pathogen using an *in situ* classifier. At this point, the implementation of the model has been substantially abstracted and does not consider the technicalities of a chemical reaction network. Our design utilizes available, labeled, whole-genome data which is input as a multivariate time-series. Translating this input into the wet-lab implementation would require that the genome sequence be input one base at a time, each base one-hot encoded as a particular oligonucleotide. Implementing the reservoir using a pool of oligonucleotide species of sufficient complexity should be possible using a system of toehold-mediated strand displacement cycles. A microfluidic open reactor could be used to control chemical species flow into and out of the reservoir. For output, different product concentrations can be monitored over time using fluorescence or a colorimetric reaction.

These initial results suggest that DNA-based reservoir computing could be a means of implementing an *in situ* viral classifier to differentiate pathogen types. This method will be transferable to other pathogens, both viral and bacterial, using available genome databases.

[1] A. Goudarzi, M. R. Lakin, and D. Stefanovic. DNA reservoir computing: A novel molecular computing approach. In D. Soloveichik and B. Yurke, editors, *DNA Computing and Molecular Programming. DNA 2013. Lecture Notes in Computer Science*, volume 8141. Springer, Cham, 2013.

[2] R. Lopez, R. Wang, and G. Seelig. A molecular multi-gene classifier for disease diagnostics. *Nature Chemistry*, 10(7):746, 2018.

[3] I. B. Yildiz, H. Jaeger, and S. J. Kiebel. Re-visiting the echo state property. *Neural Networks*, 35:1–9, 2012.

Diagonal-cut square-lattice DNA origami

Luzia Kilwing[1, 2], Bhavik Nathwani [2], Richard Guerra [2], Tim Liedl[1], William M. Shih[2]

[1] Department of Physics and Center for NanoScience (CeNS), LMU Munich, Germany
[2] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Wyss Institute for Biologically Inspired Engineering at Harvard University, Department of Cancer Biology, Dana-Farber Cancer Institute

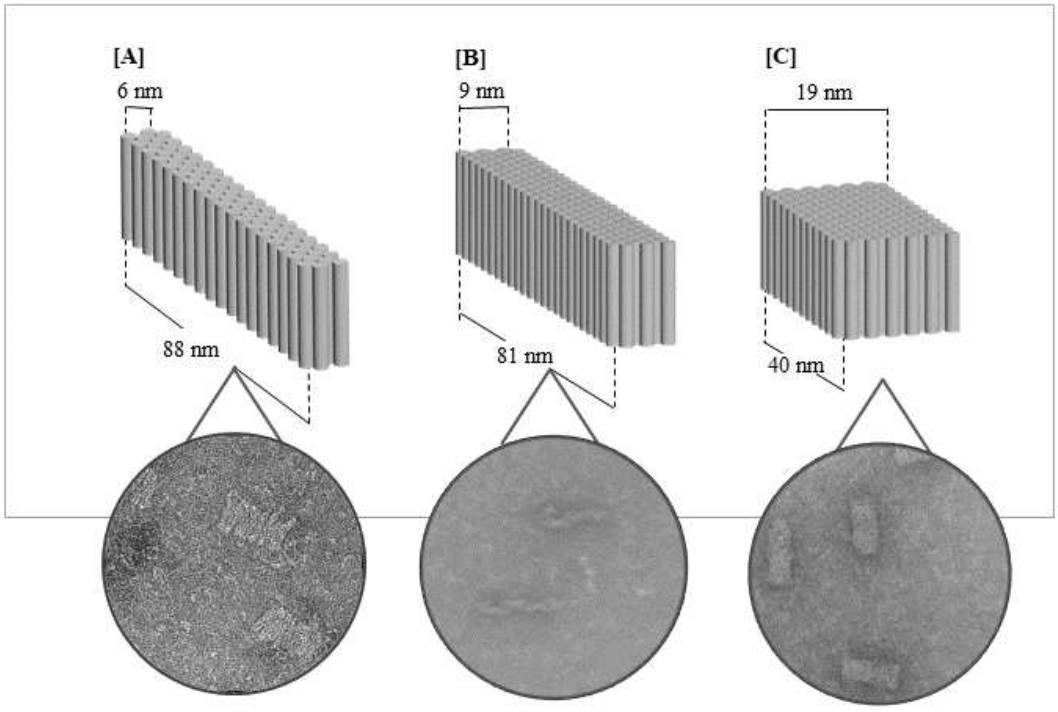
DNA origami is a popular technique to design and assemble DNA nanostructures using a ~7-kb single DNA strand as template ("scaffold") and hundreds of pre-designed oligonucleotides ("staples") to bring the scaffold into a desired shape by Watson-Crick base pairing. Versatile addressability for chemical functionalization and structural complexity on the nanoscale are some of the outstanding properties of the structures assembled with this technique. The Rothemund rectangle can display functionalities at staple-strand breakpoints with a 5 nm rhombic-lattice spacing off the sides of the constituent double helices.

Here we explore a multi-layer DNA-origami architecture with helices bundled in a diagonal-cut arrangement of a square-lattice. Display off the ends of alternating helices (e.g. even parity) yields 3.5 nm spacing in a square-lattice arrangement. In a first design **[A]**, we tested the performance of 84 helices of length 32 nm (96 bp) arranged to provide a 3×14 display. In a second and third design, we investigated whether these helices could be shortened to 17 nm (50 bp). Along the "north-south" axis in the caDNAno axial view, a pair of scaffold crossovers connects adjacent helices. However, along the "east-west" axis, only a single staple crossover connects adjacent helices.

Figure **[B]** shows 144 helices of length 17 nm (50 bp) arranged to provide a 3×24 display, while Figure **[C]** shows a 6×12 display.

Analysis of all designs with gel electrophoresis and transmission electron microscopy shows that it is possible to fold structures with a dense packing of helices suggested by Design 1A.

Design B generates highly supertwisted structures, which we attribute to the use of single staple crossovers instead of double staple crossovers. However, Design C structures are not highly supertwisted, which we attribute to the higher moment of torsional inertia that results from a smaller aspect ratio. We currently are studying how adjusting the positions along the helices of the single staple crossovers can further reduce supertwisting.



Efficient Approximation of Sequence Hybridization

David Buterez and Thomas Heinis, Imperial College London

Challenge Determining if two DNA sequences will anneal (under the right circumstances) is crucial to find a set of sequences that are pairwise orthogonal, i.e., that will not anneal, to implement either random access [3] in DNA storage or to address sequences in database operations [1]. The size of this set determines how many distinct sequences can be addressed and can thus be randomly accessed in a pool. Recent work [5] estimates that, given all possible DNA sequences of length 20, only 14'000 pairs are pairwise orthogonal, meaning that only 14'000 different sequences can be addressed (and retrieved through amplification using PCR). Testing if more sequences can be used is very costly.

Current Approach The current approach to address this challenge is to use tools like NUPACK [4] to calculate to the free energy between a pair of sequences. While this allows for a realistic prediction of the likelihood of two sequences to anneal, the calculation is prohibitively slow to perform it on a large scale. Recent work, LDA [2], uses models based on Linear Discriminant Analysis to compute an approximate result quicker. The sequences considered in their work are from biologic origin, have a length between 26 and 46 nucleotides and are a bit longer than ours.

Our Approach In our work we want to compute more accurate results quicker than LDA. We consequently approximate the NUPACK output, the yield, using machine learning models. The challenge lies in finding the best model, setting its parameters and designing the training dataset — generating a representative sample using NUPACK. Given the temperature, the sequences and whatever, the model is trained to then predict the yield.

Building a balanced training dataset — a realistic proportion the classes of pairs that anneal and pairs that do not — is challenging. We generally generate sequences at random with only a boolean parameter to exclude homopolymers and the length of the sequence as input parameters. Using purely random sequences will lead to a very low yield for the vast majority of sequences. To balance the classes we use a few random mutations on sequences with high yield to find more pairs sequences with high yield.

As opposed to other approaches, we do not consider GC content and similar constraints to keep as many degrees of freedom in the sequence generation. We focus on sequence/primer lengths between 18 and 25 nucleotides as we consider them most likely to be used as addresses.

Finally, we use a convolutional neural network (CNN) with 17 layers, predominantly convolutional, pooling and normalization layers.

Evaluation We use a 80-20 split of the dataset of 255'709 sequences for training and testing/validation. To generate the labels for the dataset (via NUPACK), we set the temperature to 57°C.

Using the CNN, we achieve a correct classification for highly annealing pairs (yield ≥ 0.8) — the class we predominantly care about — of 98.3%. The correct classification of pairs with a low yield is 94.9%. Both are substantially higher than previous methods like LDA on our data (or other machine learning methods, like random forests, we explored). The classification is very fast with 90 seconds (on a Nvidia Tesla V100 GPU) for 1'022'796 pairs of sequences — nearly two orders of magnitude faster than NUPACK and faster than LDA.

We trained the model for 57°C but also tried classification for NUPACK results at other temperatures. However, the annealing operation is very sensitive and so the prediction accuracy drops quickly (e.g., 92% for 52°C) if other temperatures are used, suggesting that a separate model has to be used for each different temperature. Still, prediction accuracy is very high for a fixed temperature.

References

1. R. Appuswamy, K. L. Brigand, P. Barbry, M. Antonini, O. Madderson, P. Freemont, J. McDonald, and T. Heinis. Oligoarchive: Using DNA in the DBMS storage hierarchy. In *CIDR 2019, 9th Biennial Conference on Innovative Data Systems Research.*
2. B. J. Beliveau, J. Y. Kishi, G. Nir, H. M. Sasaki, S. K. Saka, S. C. Nguyen, C.-t. Wu, and P. Yin. Oligominer provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proceedings of the National Academy of Sciences*, 115(10), 2018.
3. J. Bornholt, L. Ceze, R. Lopez, G. Seelig, D. M. Carmean, K. Strauss, L. Ceze, G. Seelig, K. Strauss, R. Lopez, G. Seelig, D. M. Carmean, and K. Strauss. A DNA-Based Archival Storage System. *ACM SIGOPS Operating Systems Review*, 2016.
4. R. Dirks, J. Bois, J. Schaeffer, E. Winfree, and N. Pierce. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Review*, 49(1):65–88, 2007.
5. L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. N. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, and K. Strauss. Random access in large-scale dna data storage. *Nature Biotechnology*, 36:242 EP –, 02 2018.

Encoding Information in Primers

Jeremy Ling and Thomas Heinis, Imperial College London

Challenge Primer design is crucial for applications related to DNA computing and DNA storage. Essentially, primers need to be designed such that they are specific enough to anneal to the correct target, e.g., for PCR amplification in support of random access in DNA storage [4] or to address sequences in support of database operations [1] and thus to solve combinatorial problems. They should not anneal to other, similar ones as this leads to amplification of the wrong sequences or to incorrect results when solving combinatorial problems (or specifically to spurious results due to the imprecise molecular cloning techniques when performing database operations [1]).

Current approaches typically compute a set of orthogonal primers sufficiently big such that each DNA sequence or database record (or group of either) can be addressed or retrieved individually through PCR. Each address (e.g., file name or primary key of a database record) is then mapped to a primer [2]. This mapping approach is simple but (a) additionally requires storing the mapping and (b) wastes precious space on the DNA sequence which cannot be used to store other information. Ideally — and that is the goal of our approach — primers themselves encode information to make best use of the space used for the address.

Our Approach The goal of such an encoding must be that similar addresses (or similar values in the case of database records) are encoded such that their DNA sequences are substantially different. To accomplish this, we make use of the avalanche property — small differences in input value lead to considerable differences in the computed checksum — of a set of cryptographic functions. Hence, if the values encoded are similar, their checksums are considerably different, thanks to the avalanche property.

Specifically, we use SHA3 as cryptographic function to compute the checksum of the address/value. We further use a simple encoding from binary to nucleotides proposed by Church [3] and encode both, address/value and checksum.

To then make similar address/value encodings substantially different, we split the encoding of a address/value a as well as the checksum c separately into subsets (a_1, a_2, a_3, ... and c_1, c_2, c_3, ...) and interleave one with the other (resulting in a_1, c_1, a_2, c_2, a_3, c_3, ...). Figure 1 illustrates the process and the result.

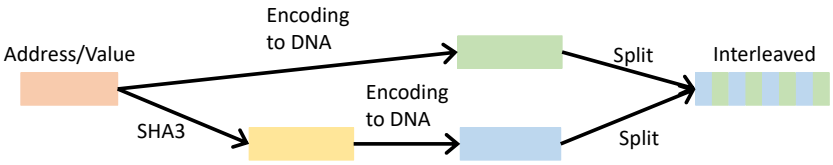


Fig. 1. Computing the interleaved address encoding.

Using a cryptographic hash function has the additional benefit that it can be used for error detection as well: the interleaved bits can be combined into the checksum which can be compared to the checksum computed on the decoded value.

Results We have used the encoding to address database records to perform in vitro database operations [1]. More precisely, we have encoded the primary keys of 1'600 records (along with the records themselves) to perform database select operations, i.e., retrieve a specific record encoded as DNA sequence from the pool via PCR. Our experiments show that we can retrieve individual sequences/records based on their primary key despite them having similar (indeed consecutive) primary keys. The fact that we are able to do so demonstrates that the primers are sufficiently different.

We plan to develop the idea further and develop cryptographic hash functions which not only allow to detect errors but to also correct errors.

References

1. R. Appuswamy, K. L. Brigand, P. Barbry, M. Antonini, O. Madderson, P. Freemont, J. McDonald, and T. Heinis. Oligoarchive: Using DNA in the DBMS storage hierarchy. In *CIDR 2019, 9th Biennial Conference on Innovative Data Systems Research*.
2. S. M. Hossein Tabatabaei Yazdi, R. Gabrys, and O. Milenkovic. Portable and Error-Free DNA-Based Data Storage. *Scientific Reports*, 2017.
3. H. H. Lee, R. Kalhor, N. Goela, J. Bolot, and G. M. Church. Enzymatic DNA Synthesis for Digital Information Storage. *bioRxiv*, 2018. https://www.biorxiv.org/content/10.1101/348987v1.
4. L. Organick, S. D. Ang, Y.-J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C. N. Takahashi, S. Newman, H.-Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, and K. Strauss. Random Access in Large-scale DNA Data Storage. *Nature Biotechnology*, 36:242 EP –, 02 2018.

A database of cadnano file to store design information of DNA nanostructure

Ibuki Kawamata[1], Sho Aradachi[1] and Satoshi Murata[1]

[1] *Department of Robotics, School of Engineering, Tohoku University, Japan*
Email: kawamata@molbot.mech.tohoku.ac.jp

In the field of structural DNA nanotechnology, DNA duplex is used as a build block to assemble nano-scale structures with complex geometry [1]. In the past few decades, various kind of DNA nanostructures has been demonstrated with a potential application in molecular-scale robotics, smart drug delivery systems, and so on [2]. To rationally design such DNA nanostructures, several softwares have been developed such as caDNAno [3], Tiamat [4], SARSE [5], vHelix [6], and DAEDALUS [7]. Although those softwares are available online, raw design information (e.g. caDNAno file) is not usually provided together with a paper publication. The situation causes an underlying problem that one has to pay a great effort to reuse the DNA nanostructures of the others. Getting rid of the obstacle may enable us to engineer a new DNA system on top of well-qualified DNA nanostructures. On the other hand, in the field of structural molecular biology and synthetic biology, atomic structure of biomolecule and gene sequence is successfully stored in databases such as protein data bank (PDB) [8] and standardized gene parts library of BioBrick [9], respectively.
Here, we propose a database that stores the design information of DNA nanostructures, especially targeting the caDNAno format (Fig. 1). In the currently developed trial version, the input to the database is a caDNAno file which is converted to several useful formats such as PDB file, rendered images, DNA sequence table etc. It is also possible to add information such as digital object identifier (DOI) to identify the publication about the structure. Moreover, we prepare some standard objects such as six helix bundles, rectangles, and cuboids, which are listed in the database. From the web-based interface, users can upload a caDNAno file and access to the data for the purpose of lab experiment and/or computer simulations such as CanDo [10] and oxDNA [11]. At the poster presentation, we would like to discuss further requirements and have some feedbacks of the database to improve the user experience of the service. In the future, the database may serve as a catalog of DNA objects, which will accelerate the engineering process of novel and complex DNA nanostructures.

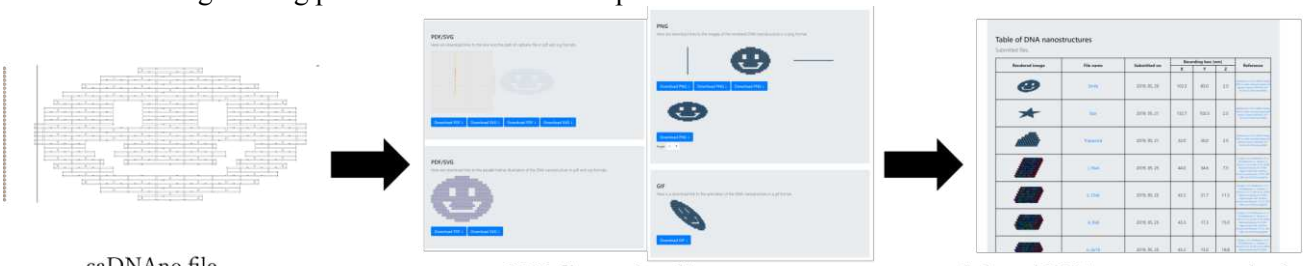


Figure 1: The schematic figure of the DNA nanostructure database. When the caDNAno file is uploaded, it is converted to several file formats, which will be listed in the database.

References

[1] N. C. Seeman, Molecular Biotechnology. 2007, **37**, 246—257.
[2] N. C. Seeman and H, F. Sleiman, Nature Reviews Materials. 2017, **3**, 17068.
[3] S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih, Nucleic Acids Research. 2009, **37**, 5001–5006.
[4] S. Williams, K. Lund, C. LinPeter, W. Lindsay, and H. Yan, Lecture Notes in Computer Science. 2008, **5347**, 99—101.
[5] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, A. Lind-Thomsen, W. Mamdouh, K. V. Gothelf, F. Besenbacher and J. Kjems, ACS Nano. 2008, **2**, 1213—1218.
[6] E. Benson, A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen, and B. Högberg, Nature. 2015, **523**, 441—444.
[7] R. Veneziano, S. Ratanalert, K. Zhang, F. Zhang, H. Yan, W. Chiu, and M. Bathe, Science. 2016, **352**, 1534.
[8] J. L. Sussman, D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Rittera, and E. E. Abola, Acta Crystallographica Section D. 1998, **54**, 1078—1084.
[9] T. Knight, MIT Artificial Intelligence Laboratory. 2003.
[10] D.-N. Kim, F. Kilchherr, H. Dietz, and M. Bathe, Nucleic Acids Research. 2012, **40**, 2862—2868
[11] T.E. Ouldridge, A.A. Louis, and J.P.K. Doye, Physical Review Letters. 2010, **104**, 178101.

Boolean Circuits via Thermodynamic Binding Networks: From Theory to Test Tube

Boya Wang[#], Cameron Chalk[#], and David Soloveichik[*]

Department of Electrical and Computer Engineering, the University of Texas at Austin

Identifying and programming equilibria is challenging in many systems. Thermodynamic binding networks (TBNs) [1] simplify analysis of energy landscapes by coarsely modeling monomers in a well-mixed solution under a specific experimental regime with two main assumptions. First and foremost, bonds are strongly favored; then, having more separate complexes is more favorable than having fewer (maximizing configurational entropy). An argument can be made that in this regime, a configuration which is maximally bonded (*saturated*) and has the largest number of free complexes among saturated configurations is in high yield at equilibrium. We call such a configuration *stable*. It has been shown that a TBN can assemble arbitrary Boolean circuits (Figure 1), where any stable configuration has the correct circuit [2]. The TBN model has had no experimental verification so far.

We aim to experimentally verify that the TBN model can be used to design DNA systems with desired equilibrium. We start from a single NAND gate, which has the property of functional completeness. With one set of gate strands, we show that changing the input strand significantly changes the signal at equilibrium to favor the correct output. This result serves as the foundation for implementing larger TBN circuits in the future. Additionally, techniques here will prove useful in implementing other TBN constructions with DNA.

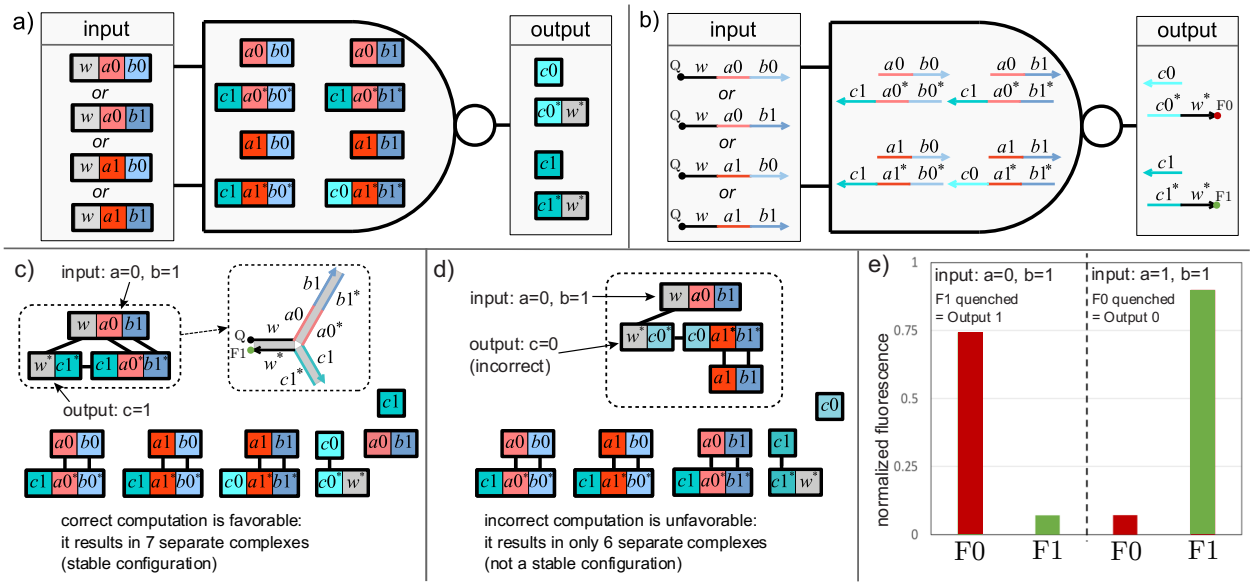


Figure 1: **a)** The TBN construction for a NAND gate, c = **NAND**(a, b). See [2] for the general construction of arbitrary Boolean circuits. **b)** The design of the DNA implementation of the TBN NAND gate. The input has a quencher Q attached, and each output strand has one of two fluorophores, F0 or F1. Which fluorophore is quenched is a proxy for which output is in the same complex as the input. **c)** A configuration of the correct computation for the input monomer *a=0, b=1*: the output monomer *c=1* is in the same complex as the input. **d)** A configuration of an incorrect computation with the output *c=0* bound to input *a=0, b=1*. **e)** Fluorescence signals of the DNA implementation with different inputs at thermodynamic equilibrium.

References

[1] Doty D., Rogers T. A., Soloveichik D., Thachuk C., Woods D. (2017) Thermodynamic Binding Networks. In: DNA Computing and Molecular Programming 23. DNA23. Lecture Notes in Computer Science, vol 10467. Springer, Cham

[2] Chalk C., Hendricks J., Patitz M.J., Sharp M. (2018) Thermodynamically Favorable Computation via Tile Self-assembly. In: Unconventional Computation and Natural Computation. UCNC 2018. Lecture Notes in Computer Science, vol 10867. Springer, Cham

---

[#] *Contributed equally*

Design of controlled promiscuous DNA interaction networks

Joseph Berleant, Kristin Sheridan, Mark Bathe
Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Currently, the successful operation of systems of synthetically designed biomolecules depends on the researcher's ability to design biomolecules that interact strongly with intended species while simultaneously avoiding cross-talk with other system subcomponents. However, naturally occurring biological systems, such as ligand-receptor binding networks, often differ from synthetically designed systems in that each ligand interacts with many different receptors and each receptor with many ligands. This interaction network structure is referred to as promiscuous because of the number of interactions possible by each molecule. While ligand-receptor promiscuity has been shown to bestow computational power difficult to achieve through purely orthogonal components, little research has been done into how to design similar synthetic biological systems. It is an open question whether biomolecules can be designed to exhibit the controlled promiscuity of some natural systems. In addition, it is unknown whether controlled promiscuity could be applied to produce biological circuits with novel behaviors, to simplify existing circuits, or to generate more complex self-assembled nanostructures.

We have developed a novel method for designing sets of DNA sequences with promiscuous, controllable relative binding affinities with each of the complements of these sequences. Given a promiscuous design specification, DNA sequences are designed by assigning a binary string to each DNA sequence such that the Hamming distance between pairs of strings corresponds to their desired binding affinity. This binary encoding is performed with an isometric graph embedding into a hypercube graph (i.e. a power of K_2) representing the space of possible binary strings. Each binary string generates a unique DNA sequence by associating each bit with a mutation site on an initial DNA sequence (Figure 1a). Using this method, sets of 10-20 DNA sequences have been designed with specifications compatible with various simple and complex graph architectures (e.g. linear and cyclic graphs). Experimental validation of these designs is currently ongoing. This algorithm is the first systematic rational design technique for designing sets of promiscuously interacting sequences, and can be easily generalized to embeddings into alternative destination graphs that more fully represent the available sequence design space. In addition, it provides a natural extension to existing strand-displacement systems, in which intruder and incumbent domains have variable, controlled affinity to a bottom strand (Figure 1b).

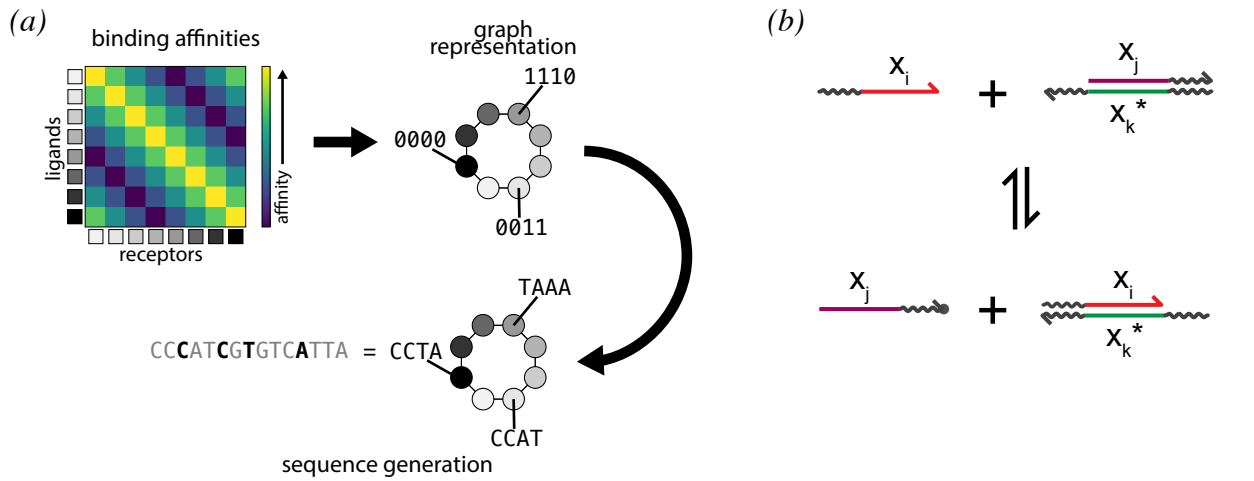


Figure 1. (a) Overview of sequence design algorithm. (b) Example strand-displacement mechanism utilizing promiscuous interactions of x_k with intruder domain x_i and incumbent domain x_j.

Quantifying effects of small changes on DNA origami nanotubes using a coarse-grained model

Domen Prešern[1,*] and *Jonathan P. K. Doye*[1,†]

[1]*Physical and Theoretical Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QZ, United Kingdom*

Effects of small changes in DNA origami designs on the global structure have long been explored (e.g. [1]), with new origami designs regularly emerging. A novel class of pleated nanotubes was recently reported by Berengut et al. [2], designed with control of global structure in mind. Simulations with a coarse-grained model of DNA, oxDNA [3, 4], were performed on a large set of closely related nanotubes. We quantified the variation of inner and outer nanotube diameters along the axes of the nanotubes, as well as the range of fluctuation in diameters and length of the nanotube and internal angles between helices due to thermal forces, depending on small differences in nanotube architecture. The degree of shear-induced internal stress in various architectures was also investigated.

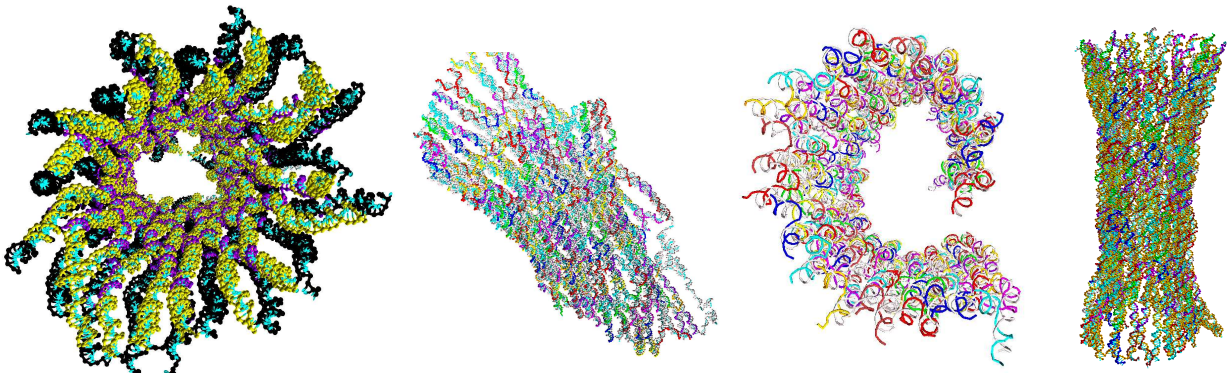


FIG. 1: Examples of simulated systems illustrating variations in shear and diameter.

* domen.presern@chem.ox.ac.uk

† jonathan.doye@chem.ox.ac.uk

[1] H. Dietz, S. M. Douglas, and W. M. Shih, Science **325**, 725 (2009), ISSN 0036-8075, URL https://science.sciencemag.org/content/325/5941/725.

[2] J. F. Berengut, J. Ruan, A. Kawamoto, and L. K. Lee, bioRxiv *preprint* (2019), URL https://www.biorxiv.org/content/early/2019/01/30/534792.

[3] B. E. K. Snodin, F. Randisi, M. Mosayebi, P. Šulc, J. S. Schreck, F. Romano, T. E. Ouldridge, R. Tsukanov, E. Nir, A. A. Louis, et al., J. Chem. Phys. **142**, 234901 (2015), URL https://aip.scitation.org/doi/10.1063/1.4921957.

[4] URL https://dna.physics.ox.ac.uk.

Modeling nucleation for self-assembled structures with uneven concentrations

Jackson O'Brien, Constantine Glen Evans, Erik Winfree, and Arvind Murugan

Introduction. For single-component self-assembly systems, nucleation theory predicts a temperature dependent critical nucleation seed size from a simple free energy analysis and a nucleation rate from an Arrhenius' approximation, using this critical nucleus as the rate-limiting step. But this classical theory of nucleation is not applicable if concentrations or binding energies vary within a structure, such as in a multi-component self-assembled structure. While such uneven concentrations do not arise for classically studied crystals made of only one or two kinds of repeating components, uneven concentrations are relevant for assemblies with many different species, such as DNA self-assembly of structures with uniquely addressable components [1].

In particular, our recent work has tried to build a pattern recognition system that classifies uneven concentrations of single-stranded DNA tiles into one of three categories through a process of self-assembly. In these systems, multiple competing structures attempt to self-assemble from the same components (multifarious self-assembly [2, 3]); the nucleation rates of some are higher because of uneven concentration distributions across each structure, and this leads to pattern recognition mediated by the physics of nucleation.

Here, we develop a modeling framework to quantitatively understand and predict nucleation rates and nucleation pathways in assemblies with uneven concentrations of heterogeneous components. This framework is not completely general, as it requires a unique tile choice at each location in the structure. However, our results and approach should be broadly applicable to many tile systems.

We propose and discuss two distinct computational models. Using these two approaches, we will compare our results to classical nucleation theory as well as experiments using uneven concentrations, estimate relative nucleation rates of competing structures in multifarious systems, and use our computational results to fine-tune multifarious pattern recognition system design.

Brief description of models. The first computational approach begins by first coarse-graining our growing crystal into a rectangular structure state-space. In this space, states will be uniquely identified by their length, width, and location within a given structure. Transitions between such states are mediated by the addition of facets with tile number corresponding to either a length or width of the parent structure. Our free energy landscape is found by calculating the free energy of each rectangular structure as well as calculating the kinetic barrier incurred during each facet nucleation process. We utilize dynamic programming for calculating kinetic barriers in this facet nucleation process as well as for the entire growth free energy landscape. By calculating a partition function over the states that precede these "critical" kinetic barrier states, we can estimate the nucleation rate for each barrier using an Arrhenius' approximation. This algorithm calculates global features of the free energy landscape but loses information about many potential growth states as it only considers rectangular assemblies of tiles.

In a complementary approach, we consider all possible starting positions within a structure. When an there are no favorable additions, a probabilistic decision is made weighted by the free-energy change of each possible addition. When favorable additions are available, the minimal free-energetic single tile addition is taken at each step. In this way, we build up a sampling of trajectories beginning at each tile location in a given structure. Along a given trajectory, the maximum free energy state on the path to complete growth is stored as "critical". This list of states can be used to estimate relative nucleation rates between structures and can be directly compared to the critical states found using the other algorithm. This algorithm uses local rules to calculate probable growth trajectories but inherently loses some information about sub-optimal paths due to finite sampling.

Outlook These two methods provide distinct approaches to estimate relative nucleation rates for multifarious systems with uneven concentrations. In contrast to more microscopically detailed studies [4], these nucleation rate calculations will still require fitting parameters before direct comparison with experiments. However, high-level schemes such as those presented here can provide practical tools for designing and evaluating various self-assembly systems.

(1) Wei, B.; Dai, M.; Yin, P. *Nature* **2012**, *485*, 623 EP -.

(2) Murugan, A.; Zeravcic, Z.; Brenner, M. P.; Leibler, S. *Proceedings of the National Academy of Sciences* **2015**, *112*, 54–59.

(3) Zhong, W.; Schwab, D. J.; Murugan, A. *Journal of Statistical Physics* **2017**, *167*, 806–826.

(4) Fonseca, P.; Romano, F.; Schreck, J. S.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A. *The Journal of Chemical Physics* **2018**, *148*, 134910.

Reservoir Computing with Random DNA Strand Displacement Circuit Systems

Hoang Nguyen, Peter Banda, Darko Stefanovic, Christof Teuscher

Implementing a top-down chemical system relies on reasoning about the functioning of the system parts in sequential causal pathways, well isolated from each other. This rather conservative approach tries to avoid complications arising from non-linear dynamics, inherent parallelism, and concurrency of chemistry and therefore considers them adversary. Current molecular machines have limited complexity, are brittle, application-specific, and require the sets of molecules and reactions to be explicitly designed. We argue that a bottom-up approach, where the species and reactions of chemical systems are selected at random, could explore the functional landscape, its phase transitions, and dynamical regimes beyond intuition by embracing these properties.

A DNA strand displacement circuit is a popular choice for a bottom-up constructed chemistry due to Soloveichik's proof [1] of a universal approximation of mass-action driven *Chemical Reaction Networks* (CRNs), which showed that complex CRNs can be programmed using DNA-based chemistry. There are four types of species in this network: upper strand, lower strand, partial double strand, and full double strand. These DNA strands react with one another under four types of reactions: binding, displacement, influx, and efflux. Two single strands react in binding reactions, resulting in a full double or a partial double strand. A displacement reaction happens between a single strand and a partial double strand, where the single strand displaces the partial double strand into two single strands and binds to one of these strands to become a double strand.

We propose that the dynamics of a random DNA strand displacement circuit system is an ideal candidate for reservoir computing, an emerging machine learning architecture. A *Reservoir Computer* (RC) [2] consists of a fixed, randomly connected recurrent neural network, the reservoir, which acts as a set of high-dimensional filters with fading memory, and a memoryless readout layer, trained by supervised learning. RC has been shown to outperform standard machine learning algorithms, especially for temporal (time-series) tasks. Without relying on specific species/reaction design or initial concentration, the random DNA strand displacement circuit can achieve complex dynamics that translate to superior learning performance. This provides inherent non-linearity and several types of dynamical regimes that are useful for designing and building a random DNA strand displacement circuit RC.

The first problem we tackled with the random chemical reservoir consisted of learning the Hamming distance between two input bitstreams. The random DNA strand displacement circuit RC successfully learned the Hamming distance, with an average NRMSE of 0.18 ± 0.38 over different perturbations introduced to the chemistry. We also tested the random DNA strand circuit RC with two temporal learning tasks: a short-term and a long-term memory task. Goudarzi *et al.* [3] and Yahiro *et al.* [4] have successfully used deoxyribozyme oscillators to solve these temporal problems. Here we show that our random DNA RC achieves 68.2% and 60.8% better performance on these two tasks compared to the deoxyribozyme oscillator RC in Yahiro *et al.* [4]. Our model, however, relies on an 82% larger variety of chemical species. We used these tasks to demonstrate the performance and feasibility of the random chemical RC and paved the way for potential wet applications, such as detecting pathogens or gene mutations.

1. Soloveichik, D., Seelig, G., Winfree, E.: DNA as a universal substrate for chemical kinetics. Proceedings of the National Academy of Sciences 107(12), 5393–5398 (2010)

2. Schrauwen, B., Verstraeten, D. Van Campenhout, J.: An overview of reservoir computing: theory, applications, and implementations. In: 15th European Symposium on Artificial Neural Networks, pp. 471–482. (2007)

3. Goudarzi, A., Lakin, M., Stefanovic, D.: DNA Reservoir Computing: A Novel Molecular Computing Approach. In: DNA Computing and Molecular Programming, pp. 76–89. Springer International Publishing, Cham (2013)

4. Yahiro, W., Aubert-Kato, N., Hagiya, M.: A reservoir computing approach for molecular computing. The 2018 Conference on Artificial Life: A Hybrid of the European Conference on Artificial Life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)(30), 31–38 (2018)

Using Variational AutoEncoders for exploring molecular reaction networks behavior space

Erika Yamazaki, Nathanael Aubert-kato
Ochanomizu University

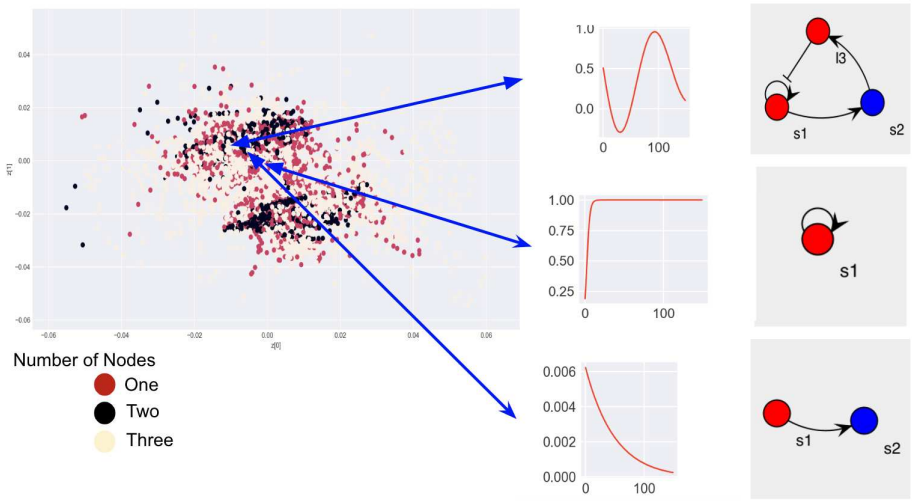
In this work, we use Variational AutoEncoder(VAE)[1] to evaluate behavior novelty in molecular reaction networks made with the PEN DNA toolbox[2]. Specifically, while the PEN toolbox can be used to design arbitrarily complex reaction networks, it is hard to explore the range of behaviors expressed by those networks. A good understanding of the characteristics of the computing paradigm is essential to guide both rational and automated design.

To estimate novelty, we want to compare reaction networks time traces. Direct comparison is difficult, due to the high dimensionality of those mathematical objects. For that reason, we combine our approach with a VAE, a specific type of neural networks used for dimensionality reduction. The network is made of two parts: an encoder that takes a high-dimensional input and returns a vector of a much lower dimensionality (latent vector) and a decoder that takes a latent vector and restores the original signal. The network is trained so that the final output is as close as possible to the original input.

We evaluated VAE using a number of latent dimensions ranging from 2 to 20. As expected, as the number of dimension increases, the accuracy of the VAE increases, however, the increased dimentionality means that exploring the latent space is harder.

We evaluated the behavior of reaction networks with 1 to 3 activation species (total reaction network size ranging from 1 to 27 signal species) by generating 1000 random networks for each size (3000 total networks). The generated time traces were then fed to the trained encoder. The resulting latent vector in the best dimension 2 VAE are shown below.

We found that most system have a similar behavior (autocatalysis or decay), but that multiple outliers were readily found. This is a promising preliminary result, showing that system algorithmic search such as Novelty Search[3] can be applied.



[1] D. P. Kingma and M. Welling. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114, 2013.

[2] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez. Programming an in vitro dna oscillator using a molecular networking strategy. Mol. Syst. Biol., 7(1):466, 2011.

[3] J. Lehman, K.O. Stanley. Abandoning objectives: evolution through the search for novelty alone, Evol. Comput., 19 (2011).

Combinatoric Catalytic Reaction Networks

William Poole
California Institute of Technology, Pasadena, CA
wpoole@caltech.edu

The tools of computer science - in particular machine learning - provide powerful frameworks to detangle the complexity which has evolved in cell and molecular biology. Chemical reaction network (CRN) models can provide insight into the structural and functional relationships of biochemical networks by considering how these networks implement various algorithms. For example, CRN models of signalling networks and genetic regulatory networks have been interpreted as neural networks [1, 2] and many abstract CRN algorithms have been proposed [3, 4]. Often, however, the specialized nature of these constructions obscures their biological relevance and raises technical difficulties with regards to synthetic implementations, especially *in vivo*.

In this poster, Combinatoric Catalytic Reaction Networks (CCRNs) are introduced as a biologically plausible model for biochemical computation at the single cell level. This model is shown to be *naturally* implementable using the components and subsystems that already exist in biology including promiscuous protein-phosphorylation networks [5] and combinatoric transcription factor regulation [2]. Importantly, the CCRN utilizes intrinsic noise in the cellular environment as a computational resource in order to perform stochastic computation. This work builds heavily off a previous work, where two versions of the CCRN were introduced and shown to be capable of Bayesian inference [6], including limited ability to generate and classify handwritten digits from the MNIST database [7]. In this work, a generalized version of the same model is presented and used to classify and generate digits from MNIST *in silico* with substantial performance gains over the previous paper.

1. E Mjolsness, DH Sharp, and J Reinitz. A connectionist model of development. *Journal of theoretical Biology*, 152(4):429–453, 1991.
2. NE Buchler, U Gerland, and T Hwa. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Sciences*, 100(9):5136–5141, 2003.
3. J Kim, JJ Hopfield, and E Winfree. Neural Network Computation by In Vitro Transcriptional Circuits. In *Advances in Neural Information Processing Systems (NIPS)*, pages 681–688, 2004.
4. Muppirala Viswa Virinchi, Abhishek Behera, and Manoj Gopalkrishnan. A reaction network scheme which implements the em algorithm. In *International Conference on DNA Computing and Molecular Programming*, pages 189–207. Springer, 2018.
5. Gideon Schreiber and Amy E Keating. Protein binding specificity versus promiscuity. *Current opinion in structural biology*, 21(1):50–61, 2011.
6. William Poole, Andrés Ortiz-Munoz, Abhishek Behera, Nick S Jones, Thomas E Ouldridge, Erik Winfree, and Manoj Gopalkrishnan. Chemical boltzmann machines. In *International Conference on DNA-Based Computers*, pages 210–231. Springer, 2017.
7. Y LeCun, C Cortes, and CJC Burges. The MNIST database of handwritten digits, 1998.

Counting with thermodynamic binding networks*

David Haley[1] and David Doty[1]

[1]University of California, Davis

An increasing demand for chemical systems and mechanisms that implement complex computation requires that computation be efficient, not just in the time to compute a result, but also in the amount of fuel compounds used during the calculation. Many existing chemical computational systems rely on kinetic controls and driving forces; however, it is a basic result of physical chemistry that chemical systems will also be driven to a thermodynamic equilibrium. Because of this, unavoidable errors caused by thermodynamic driving forces can accumulate in the system. One way to ensure that the computation is not thermodynamically driven to error is to design a mechanism in which the states of the computation are energy neutral; that is, that they should be equivalent in energy by some measure and that the pathways between them are reversible.

Under the constraint of energy neutrality, a known result is that with space proportional to the number of steps in the computation, DSDs are Turing-universal [3]. Thus, a natural question to follow is whether such computation can be conducted in a space efficient way. A result by Thachuk and Condon [4] showed that space and energy efficient computation is possible in DNA strand displacement systems. They accomplish this by re-using "fuels"; the waste product of a given reaction is also a fuel for a later reaction [2]. In this manner, an exponential number of states can be realized without also using a proportional amount of fuel or accumulating a proportional amount of waste. However, DSD systems rely heavily on geometry and on kinetic controls, and so a natural question is whether or not such a result relies directly on the geometry and kinetics of the system.

To address this question, we will look at a kinetic variant of the thermodynamic binding network (TBN) model [1], which considers only the energy barriers that arise from the number of bonds formed/broken and the entropic effect of colocating separate complexes at given steps in a reaction sequence. We are interested in the question if such a "recycling" method [2] is applicable in the kTBN model, and will answer this in the affirmative by constructing an n bit binary (Gray code) counter, which advances as a chain through the energy wells of 2^n unique stable states, while using $\Theta(n)$ monomers and $\Theta(n^2)$ domains.

This result is interesting because the results do not rely upon the specific nature of the implementing substrate, and so it suggests that the same behavior could be extended to a wide variety of chemical implementations.

References

[1] Keenan Breik, Cameron Chalk, David Doty, David Haley, and David Soloveichik. Programming substrate-independent kinetic barriers with thermodynamic binding networks. In *International Conference on Computational Methods in Systems Biology*, pages 203–219. Springer, 2018.

[2] Anne Condon, Alan J Hu, Ján Maňuch, and Chris Thachuk. Less haste, less waste: on recycling and its limits in strand displacement systems. *Interface focus*, 2(4):512–521, 2012.

[3] Lulu Qian, David Soloveichik, and Erik Winfree. Efficient turing-universal computation with dna polymers. In *International Workshop on DNA-Based Computers*, pages 123–140. Springer, 2010.

[4] Chris Thachuk and Anne Condon. Space and energy efficient computation with dna strand displacement systems. In *International Workshop on DNA-Based Computers*, pages 135–149. Springer, 2012.

*Authors supported by NSF grants 1619343 and 1844976.

Multidimensional Data Organization and Random Access in DNA Storage Systems

Xin Song[1,2,3*], Shalin Shah[1,3], John Reif[1,3]

[1]Department of Electrical and Computer Engineering
[2]Department of Biomedical Engineering
[3]Department of Computer Science

Duke University, Durham, NC 27708, USA

[*]E-mail: xin.song@duke.edu

Abstract

With impressive density and coding capacity, DNA is a promising medium for building long-lasting data storage at the molecular level. In recent DNA-based storage systems, data retrieval such as *random access* typically relies on a large library of orthogonal PCR primers. Although several algorithms have been proposed to automate the design of pooled primers, it still remains challenging to scale up. In this work, we combine the nested and semi-nested PCR techniques to enforce virtual data organization in large DNA storage systems. We strategically leverage forward/reverse primer pairs from the same or different address layers to virtually specify and maintain data retrievals in the form of rows, columns, tables, and blocks with respect to the original storage pool. This architecture enables various random-access patterns that DNA database designers could utilize and tailor to the underlying data relations of the storage content. This design alleviates the bottleneck in primer design as it only requires $4 * n$ unique primers to index n^4 data entries. Specific data subsets or individual datum from the multidimensional storage can be selectively retrieved with at most two primer pairs. The proposed architecture can be economically scaled up to organize and address very large DNA databases.

Keywords: Nucleic acid database, molecular storage, hierarchical memory, nested PCR, random access