Biochemical Controller Made From DNA

Yuan-Jyue Chen

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Georg Seelig, Chair

Eric Klavins

Jesse Zalatan

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Biochemical Controller Made From DNA

Yuan-Jyue Chen

Chair of the Supervisory Committee:
Associate Professor Georg Seelig
Electrical Engineering and Computer Science & Engineering

The potential of robots operating at a molecular or cellular scale is only limited by the imagination — for instance, nanorobots could navigate the bloodstream, identify a tumor and eliminate it cell by cell resulting in cancer treatment with minimal side effects. To perform such complex tasks, nanorobots need sensors for detecting their environment, actuators that allow them to move through their environment and embedded control circuits that convert sensor information to motor activity. In this thesis we focus on developing systematic design strategies for the *de novo* construction of embedded molecular controllers with DNA nanotechnology (introduced in Chapter 1).

To systematically engineer DNA computing systems, we developed a new class of programmable DNA circuitry that, in principle, can implement any behavior captured by chemical reaction networks (CRNs) (Chapter 2). Although CRNs have been widely used as a framework for describing and modeling the time evolution of chemical systems, we were the first to show that CRNs can also serve as a prescriptive language for specifying many complex computations, including oscillations, memory, and distributed algorithms. We thus treated CRNs as a programming language, allowing us to design our DNA circuitry while abstracting away from the molecular details. To demonstrate our approach experimentally, we constructed DNA circuits to implement a CRN that embodies, at the molecular level, an algorithm used in distributed control systems for achieving consensus between multiple

agents.

Having made significant progress in the engineering of DNA computing systems *in vitro*, we next began to explore the design principles for adapting DNA circuitry to a much more complex environment, the mammalian cell (Chapter 3). Applying DNA circuitry *in vivo*, however, requires special considerations to minimize unwanted interference from host cellular activity. We showed that the use of modified RNA bases and backbones greatly enhances circuit performance in cells. Building on this breakthrough, we constructed nucleic acid-based AND and OR logic circuits and demonstrated that they function predictably and reliably within cells. Our work is a first step toward porting the rich toolbox of DNA nanotechnology into live cells.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The past several years at the University of Washington have been among the most happy and memorable of my life. I first want to thank Georg Seelig for being a wonderful advisor, who has been a patient instructor and role model, guiding me through the many times I thought I had reached a dead end, all the while showing constant support.

I thank the other members of my supervisory committee, Eric Klavins, James M. Carothers, and Jesse Zalatan. Eric gave me interesting "big-picture" questions that helped me carefully think of the prospective of our field. James together with his post-docs, Cassandra Burke and Rodrigo Correa, brought in exciting new ideas to our DNA only systems. For example, by combining DNA strand displacement circuits with aptamer based sensors, a broad-class of biosensor control circuits can be rationally designed. Lastly, Jesse asked some questions that allowed me to think what challenges need to be met before DNA nanotechnology can become competitive with other alternative technologies.

Alex Rosenberg, Ben Groves, Richard Muscat, Sergii Pochekailov, Gourab Chatterjee, Sherry Chen, Paul Samples, and Albero Carignano have been my close friends at University of Washington. Alex has inspired me to improve myself in many ways, from the scientific to the personal. Ben has been an endless source for advice and information. I have been very lucky to work with him, and we would not be able to achieve many of our projects without Ben. Rich was always so kind to provide feedback on many of my manuscripts. Sergii has been the best source for jokes, and I thank him for always providing honest opinions. Gourab and Sherry have always been good friends and were always up for discussing our projects and sharing interesting stories. And though I've only known both Paul and Alberto relatively briefly, they've both proven to be fun to work and hang out with, and have been sources of

valuable feedback.

I thank my research collaborators, Luca Cardelli, David Soloveichik, Andrew Phillips, Neil Dalchau, and Niranjan Srinivas. Luca has always been generous with his time, and has been invaluable for his insight into my often confusing experimental data. David is always a great intellectual source, and I thank him for his extensive input into our work on chemical reaction networks. Andrew and Neil are great collaborators and also good friends; I thank them for all their hard work, particularly their efficient work on model fitting. I thank Niranjan for providing camaraderie in our shared battle towards implementing dynamic behaviors using DNA circuity. It's been a wonderful experience to work with you all, and I am looking forwards to future opportunities to work together again.

I want to thank Chiara Zurla and Philip Santangelo for collaborating in our shared exploration of the use of DNA circuitry inside cells. I thank them for providing much useful input and their hard work.

I thank my other labmates, Sundipta Rao, Sifang Chen, Nick Bogard, Randolph Lopez, Cezanne Camacho, Tim Strovas, and Matt Olson. Cezanne, Matt and Tim were the founding members of our lab, and it was wonderful to work with them. Sundipta and I worked closely together on the preparation and filming of our JoVE publication, and she proved to be a generous, easygoing and valuable collaborator. Sundipta, Gourab, Sifang, and Randolph have formed a new generation of the DNA nanotechnology community, and it's always fun to brainstorm with you guys. I also thank all the members of the Seelig group for making our lab a fantastic place to work and grow. Suffice it to say that without your presence, the lab would be a much less interesting place.

I thank NSF and Taiwanese government for providing funding for my research.

Finally, my family and girlfriend deserve my special, heartfelt, thanks. I would like to thank my girlfriend Sunny Chieh Cheng, who gives me happiness everyday, has added great breadth and depth to my life and makes me feel complete. My brother has provided

me endless invaluable advice — helping me through many of life's important decisions. I thank my sister for her support and company. My mom and dad's sacrifices for my sake are immeasurable — they have always made certain that I can follow my dreams, in a way that was not always available for them. I cannot be grateful enough to have them all in my life. This thesis is dedicated to my girlfriend and my family.

Chapter 1

DYNAMIC DNA NANOTECHNOLOGY FROM THE TEST TUBE TO THE CELL

DNA nanotechnology can be broadly divided into two areas: structural and dynamic nanotechnology. Structural DNA nanotechnology is a field in which researchers rationally design the sequences so that DNA can self-assemble into two- or three-dimensional structures. In contrast, dynamic DNA nanotechnology focuses on creating functional devices with interesting non-equilibrium dynamics. Both fields of structural and dynamic DNA nanotechnology in cell-free settings have been previously reviewed. Instead, Benjamin Groves, Richard A. Muscat, Georg Seelig, and I review the recent progress of bridging DNA nanotechnology from the test tube to a much more complex environment, the mammalian cell. The version presented in this chapter is not a final manuscript, and the emphasis is on dynamic DNA nanotechnology.

1.1 Abstract.

The programmability of Watson-Crick base pairing together with decreasing synthesis cost has made DNA a widely used material for the assembly of both structural and dynamic devices. Researchers in DNA nanotechnology have rapidly scaled up the system complexity in cell-free settings in the last decade. But the most intriguing applications for DNA nanotechnology, applications that best take advantage of the small size, biocompatibility of DNA based systems, lie at the interface with biology. Here we review the recent progress of bridging DNA nanotechnology into the cell, with particular emphasis on dynamic devices. In addition, we highlight key challenges and opportunities for cellular DNA nanotechnology.

1.2 Introduction

DNA nanotechnology is a purists approach to biomolecular engineering. The field aims to create molecular structures and devices through the exclusive use of DNA as an engineering material. The well-characterized nature of DNA base-pairing provides an easy means to control DNA interactions; this sequence programmability has allowed the rational design of precisely defined structures ranging in size from nanometers to millimeters, and of molecular motors or circuits that can autonomously move or process information.

The success of DNA nanotechnology can be traced to three key ingredients. First, our quantitative understanding of DNA thermodynamics which makes it possible to reliably predict how single-stranded DNA molecules fold and interact with one another[1, 2]. Second, the rapidly falling cost and increasing quality of DNA synthesis[3]. Third, the focus on cell-free settings, where designed reaction pathways can proceed without interference from DNA and RNA processing enzymes and other confounding factors that might be encountered in cells.

Still, DNA nanotechnology has long been motivated by the goal of building smart therapeutics, drug delivery systems, tools for molecular biology and other devices that could interact with or operate within living cells[4, 5, 6, 7]. Such applications play to the obvious strength of nucleic acid nanostructures and devices, specifically their small size, biocompatibility and the straightforward manner in which they could be programmed to interact with cellular nucleic acids through hybridization. However, in order to realize such applications using tools from DNA nanotechnology, it will be necessary to bridge the gap between performing experiments in well-mixed reaction buffers and spatially structured, densely packed cellular environments.

Here, we review recent progress towards the goal of bringing DNA nanotechnology into the cell, with particular emphasis on dynamic DNA devices. We begin with a review of dynamic DNA nanotechnology in cell-free settings then move to more cell-like environments, such as fixed cells, settings that capture some, but not all the complexity of cellular environments.

Next, we discuss several recent results showing that DNA nanodevices can be programmed to interact with cell surface proteins, before reaching devices that operate inside live cells. Finally we review initial work attempting to use DNA sensors and logic gates to detect, analyze and regulate cellular RNA levels.

1.3 Cell-free dynamic DNA nanotechnology.

In contrast to structural DNA nanotechnology[8, 9, 10, 11, 12], dynamic DNA nanotechnology aims to create devices with moving parts and time-varying behaviors[13]. The field of dynamic DNA nanotechnology can be traced to multiple ancestors including Adlemans work on DNA computation and research on the directed evolution and characterization of functional nucleic acids[14]. However, Yurke and coworkers truly launched the field by demonstrating that a functional molecular motor could be rationally designed and driven through its work cycle using only hybridization and strand displacement reactions[15]. In this section, we show how such a simple mechanism can be systematically used to create various dynamic devices, including molecular motors, computational circuits, catalytic amplifiers, and reconfigurable structures. We review these important results developed in cell-free settings and point out potential applications in the cellular environment.

1.3.1 DNA strand displacement

Strand displacement is a process in which single-strand DNA hybridizes with another partially or fully complementary complex (referred to as gates), displacing one or more prehybridized strands of the gates through the process. An example of strand displacement reaction is illustrated in Fig. 1.1. The reaction initiates when the single-stranded input binds to the complex X at the complementary single stranded domain (referred to as toeholds). Following toehold binding, the three-stranded complex undergoes a "branch migration" process [24]. At the completion of branch migration, a complex Y is formed and a single-stranded output <a tb> is released. The reaction speed of strand displacement reaction can be well controlled over a factor of 10^6 by varying the strength (length and sequence composition) of toeholds[16, 17, 18, 19]. This feature has enabled the construction of DNA nanodevices with engineering control over the reaction kinetics.

Another important feature for strand displacement is that the reactions can be easily cascaded, this allows the engineering for complex autonomous systems. Taking Fig. 1.1 as an

example, the strand displacement reaction releases a single-stranded output, and this output can serve as the input for the downstream reaction. As such, multiple strand displacement reactions could be chained together to create complex reaction networks as demonstrated by the Winfree and Pierce groups[20, 21]. Owing to their simplicity, DNA strand displacement cascades have since been used widely and effectively for molecular engineering and provide the mechanism driving most dynamic DNA devices.

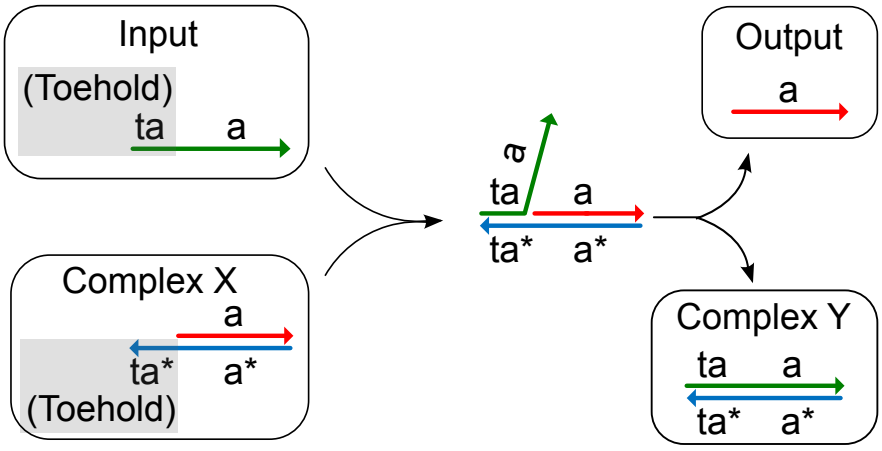


Figure 1.1: DNA strand displacement. Strand displacement mechanism. DNA strands are drawn as lines with arrows at the 3' end, and * indicates complementarity. A single-stranded DNA input (green) binds to a double-stranded DNA complex X (red-blue) via the toehold domains <ta> and <ta*>. The hybridization of the toeholds allows the input strand to displace the output strand from the complex X.

1.3.2 DNA walkers.

Inspired by motor proteins such as kinesin that moves in a step-wise manner, Sherman and Seeman as well as Shin and Pierce developed strand-displacement based DNA bipedal walkers that can move along a one-dimensional track[22, 23]. Their walkers take one step once upon the addition of a input. Gu *et al.* constructed a multi-arm DNA walker that could pick up gold nanoparticle cargos and transport to a target position[24]. While these initial designs were encouraging, their operation was not autonomous and required regular addition of fuel.

Yin *et al.* developed an autonomous walker that can move directionally[25]. However, at every step, there is a 50% chance that the movement is terminated, and thus the walking distance is intrinsically limited. Omabegho *et al.* demonstrated that another bipedal walker could walk autonomously for up to three steps[26]. Note that their walker does not have intrinsically limitation with regards to the walking distance.

These DNA walkers typically could only take 2 to 3 steps, and this is partly because of the difficulty of preparing long tracks. Recently, the DNA origami technology has become much more mature[27], which has enabled the construction of tracks with longer and more complex geometry. As a result, several recent works demonstrated that DNA walkers could move much further and even navigate through a network of tracks[28, 29, 30].

Although these artificial DNA walkers are still in their infancy, we can envision that one day DNA based molecular robots can be programmed to move directionally inside cells and perform sophisticated tasks such as picking up cargos and sorting them to specific locations.

1.3.3 Molecular circuits.

Nucleic acids circuits can serve as the brain of molecular machines and have a wide range of potential applications in biotechnology. For example, Shapiro and coworkers constructed a DNA computer that can identify specific cancers by analyzing a set of predetermined mRNA profiles[31, 6]. Their logic operation was not based on strand displacement, but relied on enzyme processing. Stojanovic and coworkers developed deoxyribozyme-based logic gates[32]

and showed that the molecular computer can play the game of "tit-for-tat[33]".

In contrast to enzyme or deoxyribozyme based computers, Seelig *et al.* demonstrated that DNA strand displacement can be used to construct enzyme-free digital logic gates[20]. Besides a complete set of logic functions (AND, OR, and NOT), Seelig *et al.* showed that leaky signal can be cleaned up using thresholds, and attenuated signal can be restored using signal amplifiers. Putting these circuits together, they showed that a multilayer circuit could generate correct digital outputs even if the input signal was corrupted by small noise.

To further scale up the circuit size, Qian and Winfree developed a standardized gate motif, allowing signal amplification and thresholding within every logic operation. Qian and Winfree showed that such a simple gate motif can be cascaded to create a four-bit square-root circuit that comprised 130 DNA strands. This design is possibly the most dramatic demonstration of the completely *de novo* design of a functional molecular system. Beyond digital logic circuits, Qian *et al.* used the same gate motif to construct a small neural network. After training *in silico*, their neural network could recall the most similar molecular events even based on incomplete information.

Even though digital abstraction has achieved incredible success in electronic circuits, one needs to be careful that synthetic molecular circuits are intrinsically analog, and their behaviors may not be well captured by digital abstraction. Soloveichik *et al.* proposed that chemical reaction networks (CRNs) may be a more appropriate "programming language" for specifying analog behaviors of molecular systems. Soloveichik *et al.* showed that DNA strand displacement circuits can be designed to, in principle, approximate any kinetics that can be mathematically expressed by CRNs[34]. Recently, we experimentally constructed a class of DNA circuitry that can implement the dynamics specified by any CRNs[35]. After testing several building block like reactions, we combined them into a network that, at a molecular level, implements an optimized algorithm for achieving consensus between multiple agents. Furthermore, Oishi and Klavins showed that any linear I/O systems can be implemented with ideal CRNs, and they demonstrated that DNA circuits can be designed to approximate the dynamics of any linear systems[36].

1.3.4 Catalytic amplifiers.

Since most nucleic acid biomarkers such as microRNA and mRNA have low copy numbers, it is very important to develop signal amplification in order to sense these biomarkers. Turberfield *et al.* were the pioneers for exploring the idea of reusable DNA catalyst[37]. In their design, the "fuel" complex was initially kinetically trapped in metastable configurations. Once a DNA input hybridized to the fuel, it opened up a fast pathway, and thus accelerated the formation of the product. Importantly, the same DNA input could participate in multiple strand-displacement reaction cycles, and result in many outputs. This DNA input was thus thought of as acting catalytically even if no covalent bonds were forming or breaking.

Several subsequent works used reactant complexes with hairpin or bulge structures[38, 39, 40], and significantly reduced the leak (uncatalysed reaction) compared to the initial design by Turberfield *et al.*. The catalytic system designed by Seelig *et al.* also allowed the sequence of output to be independent from the catalyst input [40]. This modularity made it easy to incorporate this catalytic amplification into their multilayer digital circuits[20].

Unlike previous designs driven-by the energy of base-pair formation, Zhang *et al.* designed a catalytic system that is driven by the entropy gain of additional release output[41]. The design by Zhang *et al.* was much simpler and showed faster reaction kinetics than previous designs. The authors also demonstrated that the signal gain can be further increased by cascading two catalytic amplifiers, but the untriggered reaction (leakage) was also more severe. Based on this design, Qian and Winfree engineered a simple catalytic amplifier for constructing their logic gates[42].

Rather than multi-stranded complexes, hairpins have been developed as a building block to achieve amplification. Dirk and Pierce developed a hybridization chain reaction (HCR) that could result in a double-stranded polymer up to thousands of base pairs long[21]. In their design, a single-stranded input could trigger two hairpins to hybridize to each other, resulting in a product with a single-stranded domain that is identical to the initial input. Since this product can subsequently trigger other hairpins, HCR is considered as an amplifying system.

Yin *et al.* developed a system that allows catalytically formation of multi-arm structures from hairpins[25]. Toeholds were all hidden in the stem regions of hairpins, and thus inert initially. Once upon the DNA catalyst hybridized to the hairpin, it resulted in a product with one single-stranded arm that could trigger the downstream reaction. After a series of the hybridization reactions, the DNA catalyst was released out from the multi-arm structures, which could be used again to trigger other reactions. Based on this technique, Yin *et al.* constructed a cross-catalytic system in which the product of each system could catalyze the formation of each other's, and this system showed exponential amplification kinetics.

To achieve high signal gain, circuit leakage (untriggered reaction) needs to be minimize, and this problem can at least be partly traced to sequence errors occurred during the process of chemical synthesis. To circumvent this problem, Chen *et al.* developed an enzymatic method to generate high quality hairpin gates[43]. The gate template was first cloned into a plasmid, and the clones containing sequence errors were screened out via Sanger sequencing. Large quantities of single-stranded hairpin gate could then be generated through PCR amplification and enzyme digestion. The authors showed that this method significantly reduced the circuit leakage, and a two-layer cascade of their amplifier could yield 7,000-fold signal amplification.

Compared to polymerase chain reaction (PCR), the current strand-strand-displacement-based catalysis systems shows much lower signal gain. The advantage of using strand displacement based catalysts over other enzymatic amplification is that the former is isothermal and robust across a wide range of environment conditions such as pH, salinity, and temperature, which is critical for cellular applications. With the increasing quality of DNA synthesis, the amplification gain is likely to be further boosted up, and it is possible that DNA catalyst can be used to sense molecular markers directly in living cells.

1.3.5 Reconfigurable structures.

With the precise control of geometry, size and shape, DNA nanostructures have been considered as ideal drug-delivery vehicles for fighting disease. Moreover, by combining with

strand displacement reactions or DNAzyme catalysis, DNA structures can be dynamically reconfigured, which provides new opportunities for conditional release of molecular drugs. Here we highlight one important work to illustrate this idea.

Anderson *et al.* constructed a hollow DNA box that could be opened by strand displacement with a particular DNA key[44]. This design implies that the conformation change of DNA structures can be induced by sensing particular cellular inputs, allowing the transport of drug in or out of the box in a controlled manner.

Taken together, DNA nanotechnology has resulted in a rich toolbox in cell-free settings; examples including molecular walkers that autonomously move along a track, molecular circuits that can analyze complex information, catalytic amplifiers that can sense and amplify signals or reconfigurable structures that can conditionally deliver drugs. Conversely, analysis and manipulation of molecular information in and on living cells is the one area of application where molecular devices and structures can outperform their electromechanical counterparts.

1.4 DNA nanotechnology in fixed cells.

Permeabilized cells and tissues also mimic some aspects of the cellular environment: fixed cells retain much of their structural organization, in particular the spatial distribution of mRNA and proteins. Fixed cells provide a controlled setting for visualizing the subcellular distribution of mRNAs and proteins using immunostaining or fluorescence in situ hybridization (FISH); approaches from DNA nanotechnology have already proved practically useful in increasing the sensitivity and specificity of such imaging methods. For instance, molecular probes based on a hybridization chain reaction (HCR)[21] have enabled simultaneous mapping of up to five target mRNAs within intact vertebrate embryos[45, 46]. By hybridizing a set of adaptor strands to target mRNA sequences, Choi et al. were able to controllably catalyze a polymerization reaction of two types of fluorescently labeled hairpin monomers; as a result of this catalytic hybridization reaction the fluorescent signal associated with a given mRNA is amplified and can be imaged readily using a fluorescence microscope (Fig. 1.2a).

By combining the ideas of strand displacement with single-molecule fluorescence in situ hybridization (smFISH), Raj and colleagues were able to detect single nucleotide variations within individual mRNA transcripts[47]. When performing smFISH, a collection of singly labeled DNA oligonucleotides hybridize along the target RNA transcripts in fixed cells[48]. Co-localization of multiple probes on the same transcript produces a discrete fluorescence spot that is clearly discernable using conventional fluorescence microscopy. Discrimination at the level of individual nucleotides was achieved using an additional strand displacement probe modified with a distinct fluorophore/quencher pair[47]. Probe binding through toehold-mediated strand displacement was dramatically slowed in the presence of a mismatch between the toehold and target. Co-localization of the SNV fluorophore with the transcript signal was used to verify the identity of the sequence (Fig. 1.2b).

Strand displacement in fixed cells has also been demonstrated on DNA-tagged proteins[49]. The ability to hybridize and then displace strands means that the number of protein species

that may be imaged is no longer constrained by the number of resolvable wavelengths available to the microscope (generally around four), but is only limited by the number of sequential hybridization/displacement cycles that may be performed.

DNA point accumulation for imaging in nanoscale topography (DNA-PAINT) is an approach that similarly takes advantage of the reversibility of DNA hybridization. Short fluorescently labeled DNA imager strands are used to transiently bind to complementary docking strands attached to a target[50]. The spontaneous binding and unbinding causes the fluorescence at a given point to switch between the on- and off- state, allowing individual target sites to be imaged with sub-10-nm resolution using total internal reflection microscopy. As above, the reversible nature of the DNA-PAINT means that it is not limited by the number of fluorophores; with sequential labeling allowing the reuse of fluorescent dyes. By targeting cellular proteins with antibodies conjugated to DNA docking strands DNA-PAINT was adapted to in situ 3D imaging in fixed cells[51].

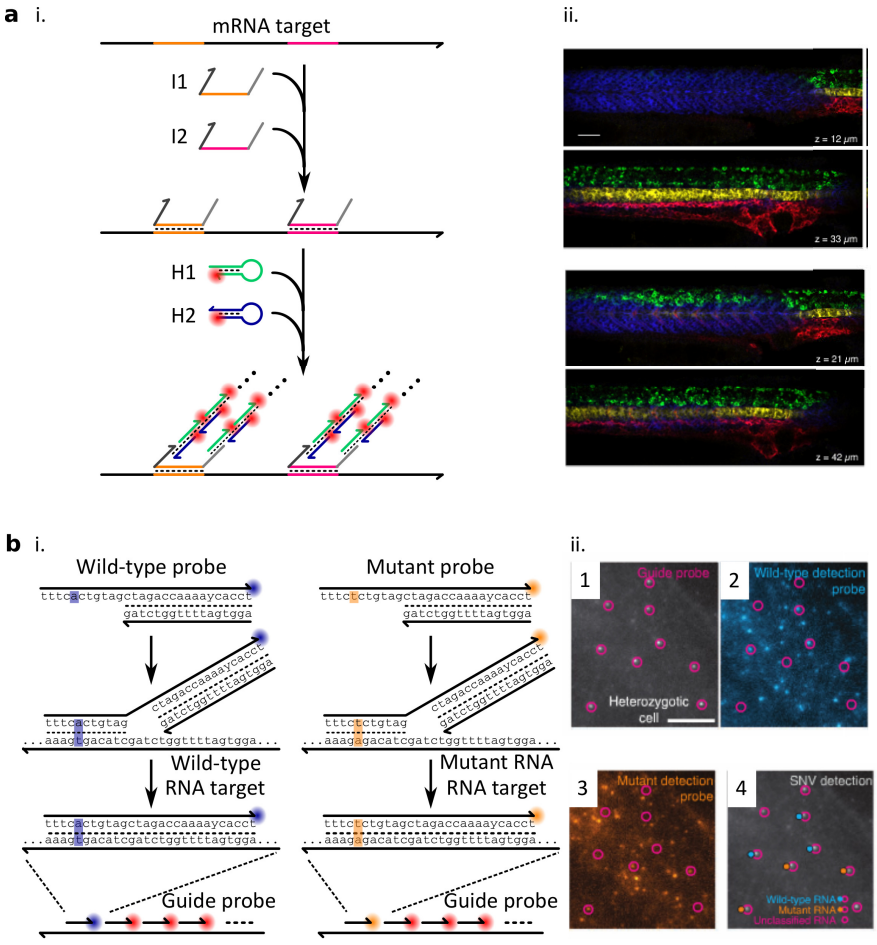


Figure 1.2: In situ imaging of mRNA in fixed cells. a, HCR FISH[46]. i. Initiator strands I1 and I2 hybridize to a target mRNA, and subsequently trigger a polymerization reaction between the two fluorescently labeled hairpin monomers H1 and H2. As a result, the target mRNA is connected to multiple fluorophores and can be visualized using fluorescence microscopy. ii. Confocal microscopy images at different z planes in a fixed zebrafish embryo. HCR probes are used to identify four different mRNAs (red: Tg(flk1:egfp); blue: tpm3; green: elevl3; yellow: ntla). Scale bar: 50 μm. **b,** Detection of a single nucleotide variation (SNV) using strand displacement probes[47]. i. Reaction mechanism. Mutant and wild-type probes compete for binding to a target mRNA. Because binding kinetics strongly depend on toehold sequence, each probe type primarily binds to the cognate mRNA. Co-localization of SNV detection probes with multiple mRNA-targeting guide probes further shows that the signal is indeed triggered off the mRNA. ii. Fluorescence micrographs of BRAF mRNA detected using guide probes (image 1), wild-type probes (image 2), mutant probes (image 3). mRNA classified as wild type or mutant (image 4). Scale bar: 5 μm.

1.5 DNA structures and devices that interact with cell surface markers.

Mammalian cells are made of a number of compartments and structures, acting as discrete vessels for biochemical reactions. The most accessible structure is the cell surface itself a lipid bilayer incorporating many surface proteins that often differentiate one cell type from another. Several recent papers demonstrated that DNA nanosystems can be designed to interact with cell surface markers; as with antibodies and aptamers[52], the most mature examples of potential DNA-based therapeutics target cell surface markers and cells in the bloodstream, targets that do not require the uptake of nanostructures into specific cells and tissues.

Douglas *et al.* created a DNA nanorobot able to deliver a molecular payload to particular cell types[53]. The payload was enclosed by a hinged origami container, which was initially held in a closed conformation, sequestering the cargo. Aptamers — DNA or RNA sequences selected to specifically bind certain proteins or even whole cells[54, 55] — provided the means for targeting the nanorobot to specific cells without the need for covalent attachment of DNA strands to antibodies. The same aptamers were also part of the locking mechanism; aptamer binding to the target protein triggered a conformational change, allowing the origami lid to open and exposing the cargo. AND logic was implemented by using combinations of two different aptamers and was used to target distinct fluorophores or a drug molecule to a subset of cells (Fig. 1.3a). Intriguingly, a similar nanorobot was also shown to be active in the bloodstream of live cockroaches, with multiple robot species performing logic operations[56].

The Tan group has also used aptamer-based logic gates to distinguish cells in mixed populations[57]. In this implementation, a scaffold is used to link multiple aptamers and a logic gate together. Binding of the aptamers to surface proteins released DNA strands that act as inputs to the logic gate. Thus, the logic gates will only be triggered if the appropriate aptamer ligands are present on the surface of the cell. Crosstalk between cells is minimal, suggesting that nearby interactions are preferential due to a higher local concentration of interacting species. The same group has recently demonstrated a more modular gate design

that allows greater number of inputs to be used, and may also be combined with the scaffold approach to improve off-target effects[58].

Stojanovic and colleagues covalently attached DNA strand displacement probes to protein antibodies to direct the probes to particular cell surface proteins[59]. Cells were labeled with one or two probes depending on which proteins were displayed on the cell surface. Following the binding stage, a trigger strand was added to activate a strand displacement cascade involving the attached probes. The output of the cascade was different depending on whether one or both probes were present, resulting in the labeling of only one of two cell types and allowing them to be distinguished (Fig. 1.3b). Although a similar outcome could be achieved by directly labeling cells with two fluorescently-tagged antibodies, this work demonstrates a more easily scalable approach for performing cell-state classification; potentially allowing many molecular markers to be analyzed in parallel and information to be summarized into an easy-to-interpret, actionable signal.

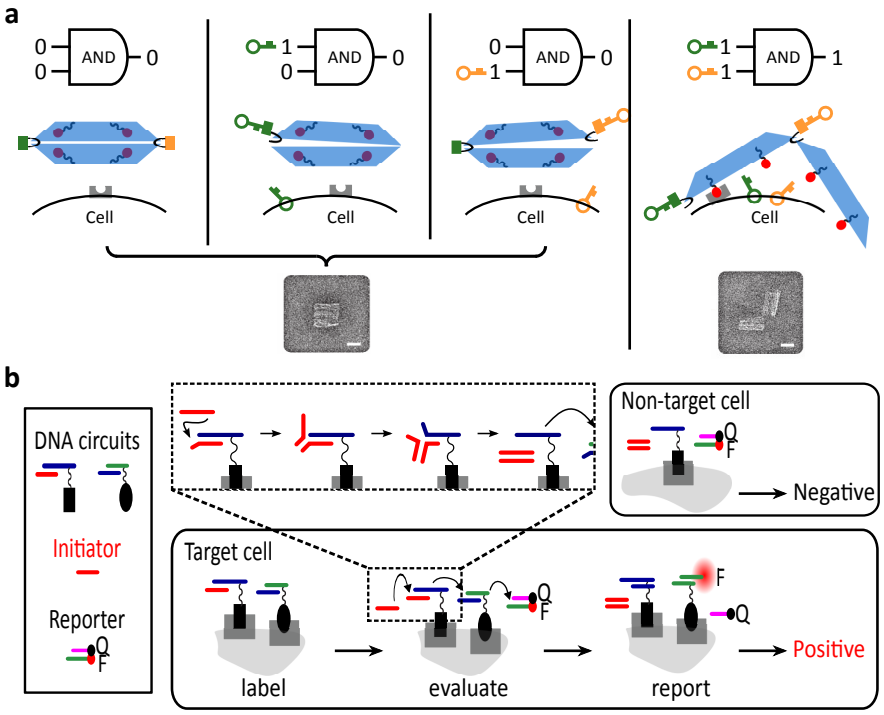


Figure 1.3: Cell surface computation. **a,** In situ cell classification through evaluating specific surface markers[59]. Cells are first coated with DNA-modified antibodies (DNA circuits; antibodies are shown as rectangles or ellipses, DNA strands as colored lines), and depending on the surface marker profiles of the cell type, either one or two gates can bind to cells. The subsequent introduction of an initiator strand I (red) triggers a series of strand displacement reactions (fully complementary strands share the same colors). A soluble reporter complex can fluorescently tag only those cells labeled with two surface-bound gates. **b,** A molecular robot for targeting a therapeutic action to specific cell types[53]. The schematic shows how a barrel-shaped nanorobot responds to specific antigens (keys) expressed on cells surfaces. The nanorobot is initially held in a closed configuration by two aptamer locks, and only when it encounters a cell displaying two matching antigens can it be opened, exposing a drug. TEM images of the closed and open states of the nanorobots are shown in the bottom.

1.6 Dynamic DNA nanodevices inside living cells.

Here, we review work on dynamic nucleic acid devices that operate within cells and respond to specific environmental cues; this includes devices that sense global environmental variables such as pH, and recent progress towards detecting specific molecular information carriers such as cellular RNA. Finally, we also discuss how the output of a nucleic acid device could result in modulation of gene expression levels and review first steps towards the construction of logic circuits for detection and analysis of multiple molecular markers.

1.6.1 Sensing the cellular environment.

Functional nucleic acids such as DNAzymes and aptamers have been used extensively within the biosensor field to detect the levels of various molecular species within cells. Here we will highlight two results that combine such sensing moieties with DNA nanomotors or structures.

Pei *et al.* constructed a set of TDNs with one or two reconfigurable edges, which allowed the TDN to change its shape in response to specific molecular signals such as protons, ATP, and mercury ions[60]. Using a FRET reporter strategy, they showed that a reconfigurable TDN changed conformation in response to intracellular ATP, demonstrating the feasibility of combining a DNA nanostructure with cellular sensors, an important property for a smart drug.

Along a similar line, Modi *et al.* used a DNA-based sensor to map the pH of endosomal pathways in living cells[61]. The design was based on Yurke's DNA tweezers — essentially two double helices connected with a flexible hinge — but incorporated an i-quadruplex structure that acts as a pH-sensitive switch, opening and closing the tweezers (Fig. 1.5a). The sensor was taken up by fly haemocytes through endocytosis and trafficked from early endosomes (pH 6) to late endosomes (pH 5.5) and finally lysosomes (pH 5). The increasingly acidic environment resulted in quantifiable fluorescence changes and thus an indirect measurement of the pH. Coupling of the sensor to the protein transferrin allowed the mapping of pH changes along a specific receptor-mediated endocytic pathway. In a follow-up paper, the same group

showed that pH-sensitive nanomachines could simultaneously track multiple pathways in the same cell[62].

1.6.2 Sensing cellular RNA.

The identity and health of a cell can often be inferred by its RNA repertoire. But the detection of specific cellular nucleic acids can be challenging because it requires nanodevices to access the cytoplasm, where most mature mRNA or miRNA are located. Furthermore, the low copy number of many RNA species may make signal amplification necessary, and the secondary structure or RNA-binding proteins can reduce the accessibility of certain sequences.

The live cell-imaging field has addressed many of these issues and has developed a number of nucleic acid probes for detecting specific mRNA sequences in live cells. Chemical and structural modifications used to improve the performance of live-cell imaging probes could also be used to enhance the intracellular performance of DNA nanodevices. Molecular beacons — stem-loop probes with a fluorophore and a quencher attached to the stem — are likely the best-studied class of probes for detection of mRNA in living cells[63]. Fluorescence is quenched when the probe is delivered but becomes unquenched when the probe hybridizes to the target mRNA. Variations on this basic design have resulted in improved performance: the addition of a longer double-stranded RNA domain or a tRNA sequence resulted in active export of the probe from the nucleus to the cytoplasm, thus facilitating interactions with mRNA[64, 65, 66] (Fig. 1.4a). Chemical modifications, in particular 2'Ome RNA, have been used to enhance probe stability by protecting against degradation by cellular nucleases[67]. As in the single-cell FISH techniques described above, co-localization of multiple probes on the same transcript can improve the signal to background ratio[64]. Multivalent probe designs such as MTRIPs in which several linear oligonucleotides are attached to a streptavidin core can also result in a stronger signal[68] (Fig. 1.4b).

Nano-flares, developed by Mirkin and co-workers, effectively provided the first example of a DNA strand displacement reaction with an RNA input in live cells (Fig. 1.4c). Nano-flares

consist of gold nanoparticles functionalized with DNA oligonucleotides complementary to an mRNA or miRNA target. Shorter fluorescently labeled oligos are hybridized to the nano-flares and are quenched by the proximity to the gold nanoparticle[69]. Binding to the target displaces the fluorescently labeled strand, resulting in an increase in fluorescence. A modified version of the nano-flare technology used an LNA-modified oligonucleotide to increase binding strength to RNA targets whilst simultaneously targeting them for degradation[70, 71]. Halo *et al.* further demonstrated that, in combination with flow cytometry, nano-flares can be used to distinguish live circulating tumor cells in the context of whole blood[72].

1.6.3 Modulating cellular RNA.

Existing non-coding nucleic acid technologies for gene-regulation, such as ASOs, ribozymes and, siRNAs can not only inform the design of cellular DNA nanosystems, but could also be integrated with DNA nanodevices as a means for controlling the cellular environment.

Afonin *et al.* demonstrated the assembly of a functional siRNA from two DNA:RNA complexes that individually did not enter the RNAi pathway in cells[73]. The reaction was initiated by hybridization of complementary single-stranded RNA overhangs present in the two inactive complexes and likely proceeded through four-way strand exchange[74]. SiRNA activity was observed in cell culture and tumor xenograft mouse models.

Activation of a regulatory response conditional upon the detection of a specific molecular marker would be even more intriguing. Benenson and colleagues took a first step in this direction and designed nucleic acid displacement circuits that interact with components of a cell lysate[75]: Detection of an RNA sequence added to the lysate triggered a strand displacement reaction leading to the creation of a functioning siRNA. Pierce and collaborators demonstrated a more general mechanism for the conditional formation of a Dicer substrate RNA also in a cell-free biochemical assay[76]. Yokobayashi and coworkers built a genetically encoded RNA hairpin system that formed a substrate for the RNAi pathway upon activation by a synthetic, exogenously delivered input oligonucleotide[77].

1.6.4 Molecular computation.

The realization of multi-input, multi-layer molecular circuits is one of the major accomplishments of dynamic DNA nanotechnology. But what unique advantages would DNA nanotechnology bring to engineering cellular biocomputers compared to alternative technologies based on synthetic gene regulatory networks? First, DNA circuits rely on components that are rationally designed at the molecular level and mechanistically simple, thus providing a high degree of control over the reaction pathway. Second, new, orthogonal components can be designed simply by changing sequence, making it easy to increase system size in a modular fashion. Third, most dynamic DNA devices have a relatively small DNA footprint compared to systems assembled from genetically encoded proteins.

The Shapiro group microinjected a DNAzyme AND gate along with microRNA-derived inputs into MCF7 breast carcinoma cells[78] (Fig. 1.5b). The gate was protected from nucleases by the addition of inverted thymidine groups to the 3' ends. Gate activation was quantified using fluorescence microscopy and, consistent with AND logic, fluorescence increased only in the presence of both inputs.

Taking it a step further, strand displacement DNA logic gates have also recently been used to detect combinations of microRNA in living cells[79]. Based on designs first demonstrated *in vitro*[20], Hemphill and Deiters used an AND gate constructed of DNA to detect the endogenous microRNAs miR-122 and miR-21 in Huh7 hepatocellular carcinoma cells[79]. Gates were delivered using standard transfection reagents and gate activation was observed only in cells that produced both input miRNA.

Together, the results reviewed in this Section suggest that the field is making rapid progress towards the design of dynamic DNA devices that can sense information in cells, analyze that information using embedded molecular control circuits and then respond to that information by effecting changes in the cell.

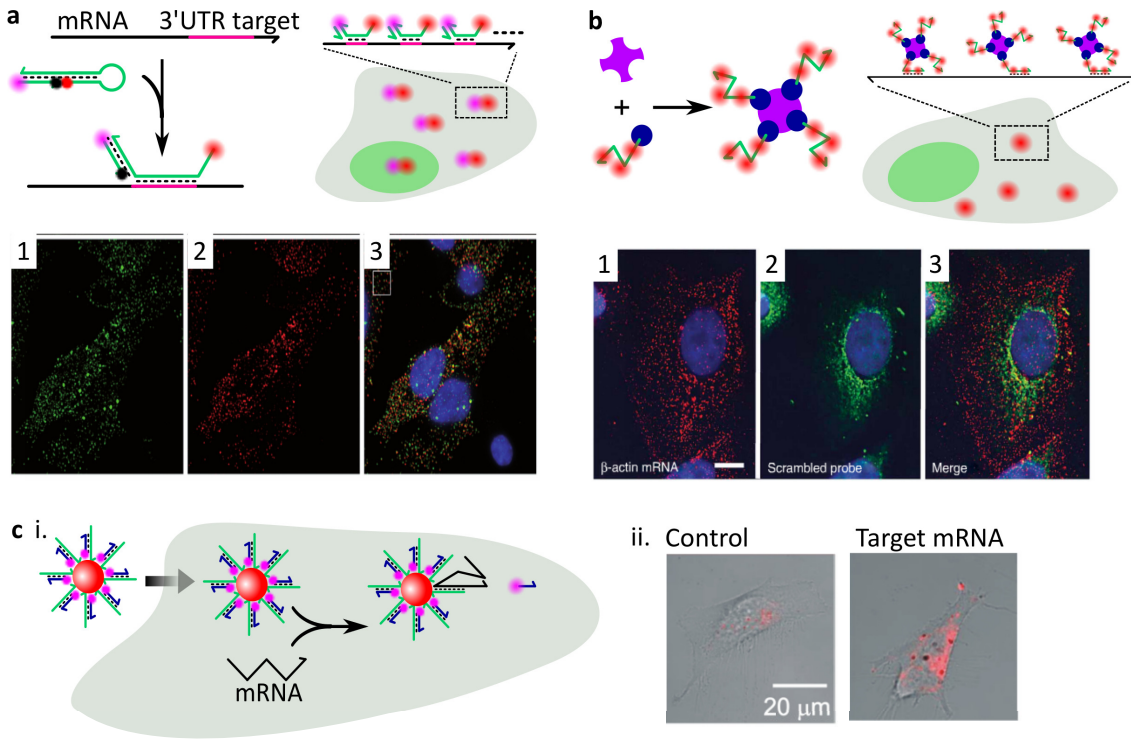


Figure 1.4: mRNA imaging in living cells. **a,** Ratiomeric biomolecular beacons (RBMBs)[66]. (Top) Binding to a target mRNA separates the reporter dye (red dot) from the quencher (black dot), resulting in high fluorescence. Multiple RBMBs can bind to the tandem repeat targets in the 3' UTR of an mRNA, enabling visualization of a single transcript in living cells. A reference dye (pink dot) is used to control for cell-to-cell variation in molecular beacon delivery. (Bottom) Fluorescence images of HT1080 cells using RBMB and FISH probes for the same mRNA. Image 1: Fish probes. Image 2: RBMB reporter dye. Image 3: A merged image that also includes nuclear DAPI stain (blue). **b,** Multiply labeled tetravalent RNA imaging probes (MTRIPs)[68]. (Top) MTRIPs consist of multiple fluorophore labeled oligonucleotides attached to streptavidin (purple). Multiple MTRIPs can be designed to hybridize to a target mRNA, making single mRNA visible in living cells. (Bottom) Deconvoluted confocal microscopy images of individual beta-actin mRNA in an A549 cell. Image 1: MTRIPs; Image 2: Scrambled probes; Image 3: A merged image that includes nuclear DAPI stain. **c,** Nanoflares[70, 71, 72]. (Left) A nanoflare contains long "capture strands" and fluorophore-labeled "flare strands," which are initially quenched by the gold nanoparticle. Target mRNAs can bind to "capture strands," displace the "flare strand" and trigger an increase in fluorescence. (Right) Confocal fluorescence microscopy images of HeLa cells treated with either control nanoflares (left image) or Survivin (target mRNA) nanoflares (right image).

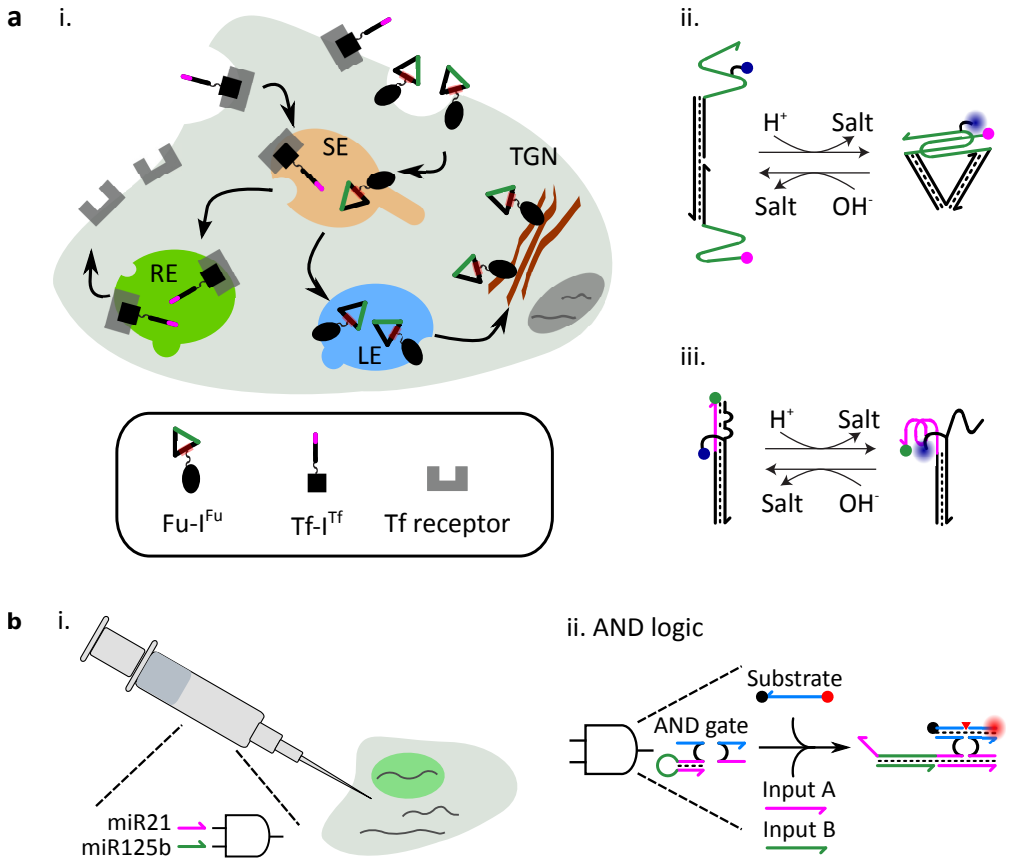


Figure 1.5: DNA nano machines and logic gates in mammalian cells. a, pH sensitive DNA nanomachines for simultaneously probing the furin (Fu) and transferrin (Tf) pathways[62]. i. A transferrin-modified DNA nanomachine Tf-I^{Tf} is confined to the transferrin pathway: it enters to sorting endosome (SE), then a recycling endosome (RE), and eventually returns to the membrane. The DNA nanomachine Fu-I^{Fu} targets the furin pathway: it enters the SE, then late endosome (LE), and eventually localizes in the trans-Golgi network (TGN). Nanomachine fluorescence is sensitive to pH which varies between different endosomal compartments. ii. pH sensitive elements of DNA nanomachines I^{fu} (green strands) and iii. pH sensitive elements of DNA nanomachines I^{Tf} (pink strands) form i-motif at low pH, resulting in high FRET between two fluorophores. **b,** A DNAzyme-based AND logic gate operates inside living cells[78]. i. Synthesized inputs with the sequences of miR-21 and mir-125b are micro-injected together with the logic AND gate. ii. Reaction mechanism: Input B first binds to the hairpin (green segment), which is then available to interact with input A to join the two components of the AND gate. The joined DNAzyme complex can then cleave the substrate, leading to high fluorescence by separating a fluorophore (red dot) from a quencher (black dot).

1.7 Outlook.

The most attractive prospect for dynamic DNA nanotechnology in cellular applications is their capacity for responding to their environment. Molecular logic and conditional (un)hiding of drug moieties could decrease side effects and increase specificity. Even the relatively simple one or two-input systems built so far have resulted in increased specificity and performance, and could be further improved with more complex multi-input logic. Diagnostic and therapeutic decisions are routinely based on the analysis of panels of multiple molecular markers, be they proteins, RNA, DNA, lipids, sugars or metabolites. For example, immunologists often need to consider large numbers of cell surface proteins to delineate all of the various cell types in a blood sample. Gene expression classifiers that reliably distinguish different tissues and disease states are typically built on measurements of tens or hundreds of different RNA species. Given the success of dynamic DNA nanotechnology in scaling up the size and reliability of molecular circuits in cell free settings, it is intriguing to think that DNA biocomputers could eventually perform complex diagnostic tasks based on the analysis of tens of molecular markers directly in living organisms.

In reality, there are many hurdles that must be overcome before DNA based therapeutics can be used in the clinic. First, the delivery of exogenous nucleic acids through systematic administration remains challenging. DNA devices need to reach the target cell before clearing by the liver and spleen. Second, potential immunostimulatory properties of exogenously delivered nucleic acids need to be investigated. Third, the cellular environment is crowded with other proteins, which could significantly affect the performance of multi-logic circuits developed *in vitro*. Finally, production cost could be a concern. While DNA circuits typically have small sizes, the cost for constructing a large nanostructure such as DNA origami is around $700[11]. Moreover, if the DNA nanostructures primarily serve as drug carriers, they have to compete with polymer materials that can cost less than $1/g. However, it seems very likely that these challenges will be overcome, given the truly unique potential of DNA nanodevices to serve as programmable, multifunctional, therapeutic systems that

could eventually rival viruses in sophistication.

Beyond diagnostic and therapeutic devices, we could imagine synthetic DNA ecosystems that integrate motors, logic, structural elements and more to control and interrogate cellular behaviors in time and space. To realize such a vision, we still need to identify broadly applicable design principles that make it easy to translate any device that works reliably in cell-free settings to the cellular environment (Fig. 1.6). New design strategies that need to be developed might include the delivery method, nucleic acid chemistry and sequence design or even different reaction mechanisms. However, given the progress that has already been made, it is quite likely that DNA nanotechnology will become a useful complement to more traditional approaches for manipulation and controlling biological information.

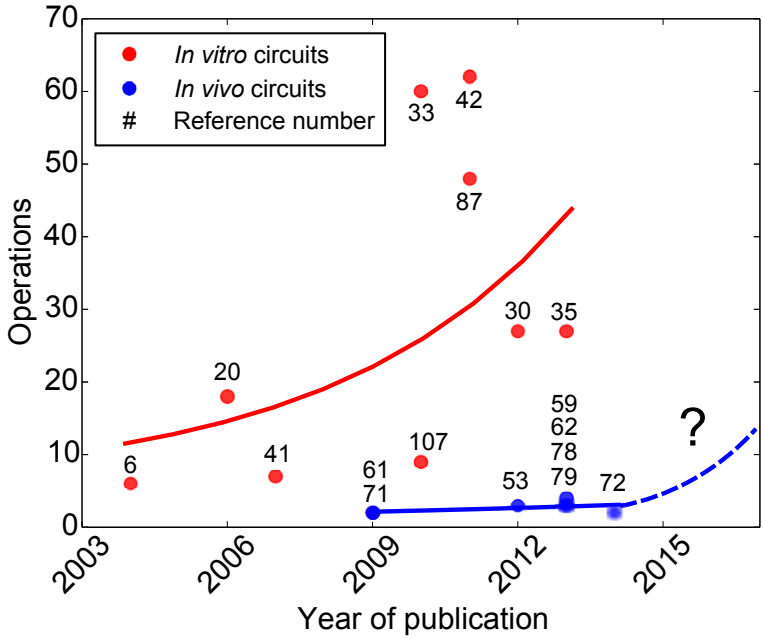


Figure 1.6: Complexity break for cellular DNA nanodevices? The complexity of cell-free DNA logic circuits and similar dynamic devices has increased by almost two orders of magnitude over the last decade. In cellular settings, only dynamic devices with maximally two or three independent operations have been demonstrated so far. This suggests that design principles for adapting dynamic DNA nanodevices to cells still need to be uncovered. Each colored dot and number represent a specific reaction network and associated publication, while trend lines are included to guide the eye. An operation is defined as a unique (sequence-specific) connection, for example a strand displacement reaction or DNAzyme cleavage event within a network. A circuit with n gates arranged in a cascade is considered to be equally complex as a circuit with n independent gates operating in parallel even though the latter is likely easier to realize experimentally. Moreover, multi-turnover catalytic reactions are weighed equally to single-step reactions, likely undercounting the complexity of the former.

Chapter 2

PROGRAMMABLE CHEMICAL CONTROLLERS MADE FROM DNA

Similar to using an electronic computer to control macroscopic, electro-mechanical robots, we aim to develop general purpose biocomputers that could serve as the brains of nanorobots. Recently we have taken a first step towards this goal and have developed a new class of programmable DNA circuitry that, in principle, can implement any behavior captured by chemical reaction networks (CRNs).

The idea that DNA strand displacement cascades can be designed to approximate any CRNs was from David Soloveichik, Georg Seelig, and Erik Winfree's theoretical proposal[34]. In the work below, I experimentally realized another class of DNA circuits that can implement arbitrary CRNs (based on Luca Cardelli's theoretical proposal[80]), and combined them into a network that embodies an algorithm for achieving consensus network between multiple agents. I also developed an enzymatic method for deriving our circuit components from plasmid DNA, which allows us to overcome the practical limitations in the length and purity of synthetic strands. Andrew Phillips and Neil Dalchau performed the modeling work. David Soloveichik made the theoretical derivation that shows our DNA circuitry can approximate the kinetics of any CRNs. Parts of this work was presented at the 17th International DNA Computing conference in Pasadena, CA, USA. This work was published in full as:

Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik and Georg Seelig, "Programmable chemical controllers made from DNA." Nature nanotechnology 8.10 (2013): 755-762.

2.1 Abstract

Biological organisms use complex molecular networks to navigate their environment and regulate internal state. Engineered molecular circuits that perform integrated sensate, computation, and actuation could open up a host of potential applications both in nanotechnology and in medicine. We report a DNA-based technology for implementing the computational core of such nano-controllers, operating in wet environments that are inherently incompatible with traditional electronics. We use the formalism of chemical reaction networks as a programming language, and our DNA architecture can, in principle, implement any behavior that can be mathematically expressed as such. Unlike logic circuits, our formulation naturally allows complex signal processing of intrinsically analog biological and chemical inputs. Controller components can be derived from plasmid DNA in high quality and quantity, supporting complex dynamics by reducing error over chemically synthesized DNA. We implement several building-block reaction types and then combine them into a network that realizes, at the molecular level, an algorithm used in distributed control systems for achieving consensus between multiple agents.

2.2 Introduction

Molecular devices have captured the imagination of chemists and engineers for at least 30 years[81]. Rationally designed active molecules include nanoparticles for the targeted delivery of drugs and imaging agents[82], or molecular motors that move along tracks and deliver cargo[83]. DNA nanotechnology[8, 13] is in a unique position among the many actively pursued strategies for constructing molecular nano-robots, demonstrating progress towards the rational design of all the required elements: sensors and amplifiers[21, 40, 37, 25, 41, 84], circuits[80, 36, 85, 86, 42, 87, 20, 34, 88, 31, 89, 90, 91], motors[28, 26, 29, 92, 15], and structures[8, 27, 93]. A rationally designed molecular robot even combined structural elements with sensing and actuation, though it lacked complex embedded control[53]. The DNA-only construction of digital logic circuits and Boolean neural networks with over a hundred rationally designed parts forms possibly the most dramatic demonstration of a systematic engineering approach to building molecular circuits[42, 87]. However, these approaches to constructing molecular information processing systems do not realize the full spectrum of analog and temporal dynamics naturally present in chemistry, and that can be harnessed to control active molecular devices.

In this work we experimentally demonstrate a design strategy for building DNA-only chemical controllers capable of being programmed to execute analog temporal dynamics. The technology is designed around a signalling protocol based on short single-stranded DNA sequences. Molecular sensors (e.g. aptamer switches) can release or expose such short sequences, and actuators (e.g. antisense drugs or ribozymes) can be triggered by them. MicroRNAs can also be used as inputs to DNA circuits[20, 79]. The control system we design sits in between, receiving inputs in the form of DNA sequences, and producing outputs in the form of other sequences (Fig. 2.1 a). The treatment of controller, sensor and actuator as independent modules has proved indispensible in other fields of engineering.

Our DNA components are, in principle, capable of realizing the entire diversity of dynamic behaviours of chemical kinetics as mathematically captured by a chemical reaction

network (CRN)[80, 34]. Although CRNs started out as a tool to understand experimental observations of elementary chemical reactions, they form a general framework to model systems with many interacting components such as gene regulatory networks, animal populations and sensor networks. CRNs can embody a wide range of digital and analog behaviours including temporal pattern generation, multi-stability and memory, Boolean logic, signal processing, control systems, or distributed algorithms[36, 94, 95, 96, 34]. Moreover, viewed as a programming language, CRNs provide a natural and intuitive formalism for delineating and reasoning about molecular interactions, without making underlying physical details explicit.

We use the familiar language of chemistry to write programs for our DNA architecture (Fig. 2.1a). The instruction A + B $\rightarrow$ C + D means that the signals A and B are transformed into the signals C and D, where A, B, C, D are DNA strands we design. The reaction is not elementary rather it is systematically compiled into a sequence of DNA strand displacement reactions. Our use of this chemical programming language is not gratuitous: A central contribution of this paper is to provide experimental evidence that our DNA architecture produces the expected stoichiometry and mass action kinetics of chemical reactions, so that our algorithms can behave similarly to what one might naively expect.

We test the major reaction classes — non-catalytic, catalytic, and autocatalytic reactions. Then we combine multiple such building blocks into a network implementing a distributed control algorithm for achieving consensus between multiple agents. While the connection between distributed computing and chemistry has been noted multiple times in the literature (e.g. Petri nets [97]), the sophistication of molecular engineering required has deterred experimental implementations. Our experiments corroborate that we can realize complex behaviours previously out of reach of synthetic molecular systems.

Among the many proposed architectures for strand displacement computation[82, 41, 84, 80, 36, 86, 87, 42, 20, 34], ours is unique in that it relies exclusively on linear, double-stranded DNA complexes (processed by nicking one of the strands)[41]. Because this structure is compatible with natural DNA, we are able to produce our computational elements in a highly pure form by bacterial cloning. Thus, we bypass the practical limitations in the

length and purity of synthetic strands.

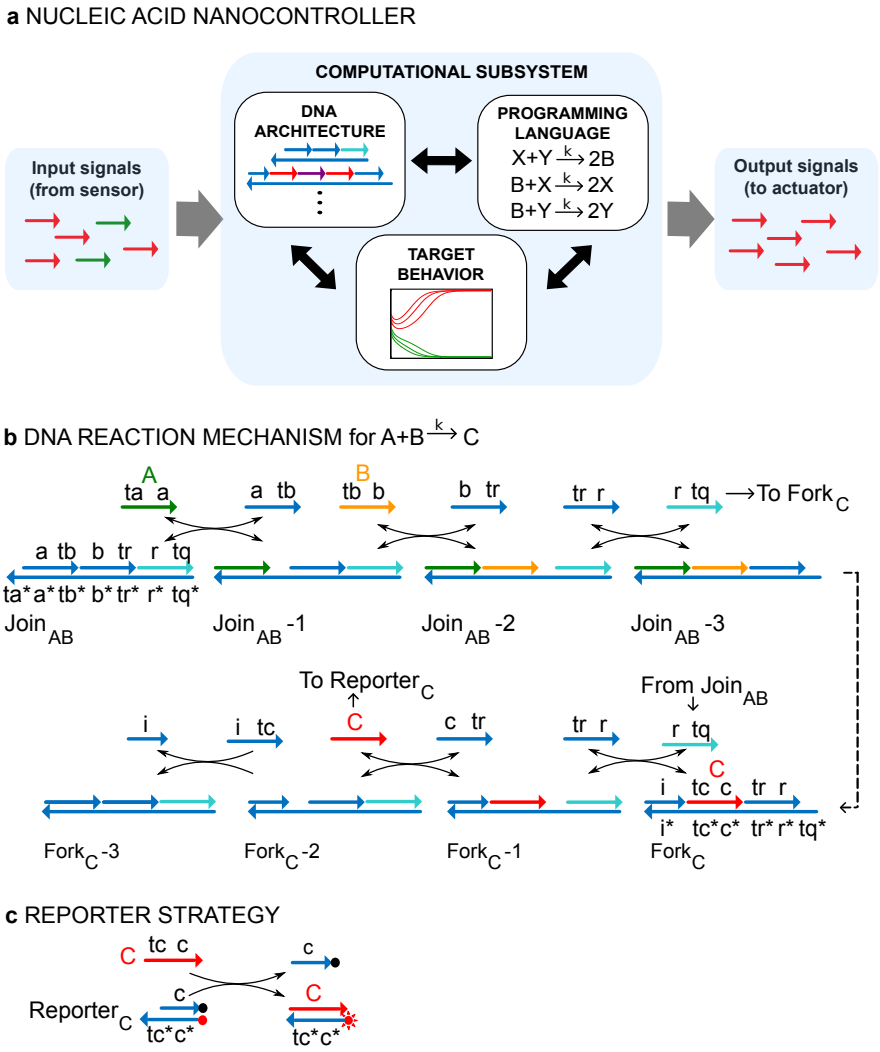


Figure 2.1: DNA realization of a formal CRN. a, A standardized signalling protocol based on short single strands of DNA enables the components of the nanocontroller to communicate with each other. The formalism of CRNs serves as a programming language that specifies the desired behaviour for the computational subsystem. The target behaviour is experimentally realized by the DNA architecture. **b,** Reaction mechanism. DNA strands are drawn as lines with arrows at the 3′ end. Functional domains are labelled with lowercase letters; * indicates WatsonCrick complement. Species A, B and C of the formal reaction are represented by DNA signal strands A (<ta a>, green), B <tb b>, orange) and C (<tc c>, red), respectively. Implementation of the bimolecular reaction A + B → C requires two multi-stranded gate complexes Join_{AB} and Fork_C as well as the auxiliary strands <tr r>, <c tr>, <i tc>. The reaction proceeds through a sequence of six stand displacement reactions where each step provides a toehold for initiation of the next one. **c,** Reporting strategy for reaction kinetics used in this paper. The reporter consists of two strands, one labeled with fluorophore (red dot), the other with a quencher (black dot). Fluorescence is quenched when fluorophore and quencher are co-localized; displacement of the quencher-labeled strand by signal C leads to an increase in fluorescence proportional to the amount of C detected.

2.3 Signal Transduction Mechanism.

We identify signals (A, B, C, ...) with single-stranded DNA molecules (signal strands, Fig. 2.1b). Nicked double-stranded DNA (ndsDNA) gate complexes mediate interactions between these signal strands with the help of additional auxiliary single-stranded species. All signal strands have the same sequence domain structure (see e.g. signal strands A (<ta a>, green), B (<tb b>, orange) and C (<tc c>, red) in 2.1b) with a short toehold domain (labels ta, tb, ...) that initiates binding to a gate, followed by a long domain (a, b, ...) that determines signal identity.

The reaction A + B → C is implemented with two gates (called Join_{AB} and Fork_C in Fig. 2.1b). The join gate consumes (and thus joins) the two signals A and B and the fork gate releases the signal C, which is initially bound to the fork gate Fork_C, and thus inactive. (The name fork gate derives from the fact that multiple signal strands can be released, as shown in later examples.) The complete triggering of a join and a fork gate — corresponding a single formal reaction A + B → C — is a cascade of strand displacement reactions in which each reaction exposes a toehold for the subsequent reaction (Fig. 2.1b). The displacing strand is either a signal strand, an auxiliary strand, or a strand previously released in the cascade (e.g. translator strand <r tq> is released by the join gate and triggers the fork gate). We use a fluorescent reporter strategy to detect specific strands and follow the reactions (Fig. 2.1c).

Each reaction is reversible until the very last displacement step involving the fork gate. The reversibility of the first step is essential to ensure stoichiometric correctness: the first formal reactant (A) should not be consumed in the absence of the second (B). Reversibility allows A to be re-released if the cascade does not complete.

The two-gate design and use of auxiliary strands ensures that all signal strands have the same domain structure and independent sequence, which guarantees composability[80]. Signal strands can thus be shared between multiple reactions to create a coupled system. Without these constraints, we can implement an individual bimolecular reaction with many

fewer species, but we would lose the ability to compose reactions into arbitrary CRNs.

2.4 Plasmid encoding of DNA gates.

The performance of strand displacement systems is currently limited by undesirable side reactions: leaks (the spontaneous firing of a reaction cascade in the absence of the intended molecular trigger), or sub-stoichiometric completion levels (unintentional sequestration of the signal leading to reduced product yield). These problems can at least in part be traced to errors in chemical DNA synthesis[98]. Biologically synthesized DNA is a useful alternative to synthetic DNA even in non-biological applications where large quantities of highly pure DNA are required[99, 43, 100].

Our gates consist entirely of nicked double stranded DNA[80], which makes them uniquely compatible with plasmid DNA as a starting material. Plasmid-derived gates have the additional advantage that they can be replicated and stored as bacterial glycerol stocks (before enzymatic processing). Gate production is detailed in Fig. 2.2a. Correct processing was tested using gel electrophoresis (Fig. 2.2). The sequence constraints imposed by the use of nicking enzymes do not limit the generality of our method. Signals can be made orthogonal to one another by designing the sequences surrounding the nicking sites to be different. All data shown in the paper was collected with plasmid-derived ndsDNA gates except where otherwise indicated. Externally added signal and auxiliary strands, as well as the reporter gates used for following reaction kinetics were chemically synthesized.

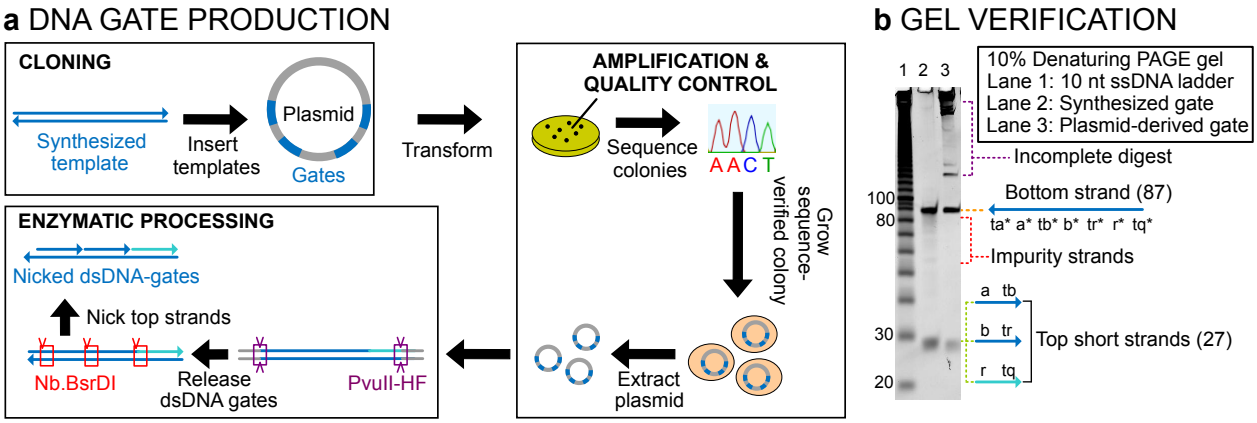


Figure 2.2: DNA gate production. a, Highly pure ndsDNA gates can be produced from plasmid DNA. Multiple copies of a double-stranded ndsDNA gate template are inserted into a plasmid and transformed into E. coli cells. Clones are picked and plasmid sequence is verified. A clonal population is grown up, and plasmid DNA is extracted using standard molecular techniques. Finally, the restriction enzyme PvuII-HF is used to release the gate from the plasmid, and the nicking enzyme Nb.BsrDI is used to generate nicks in the top strand. **b,** Analysis by 10% denaturing polyacrylamide gel electrophoresis (PAGE) of the enzymatically processed gate. The long bottom strand (87 mer) and short top strands (27 mer) are visible on the D-gel.

2.5 Testing fundamental reaction types.

The modular nature of our design makes it easy to create reactions with multiple products of unconstrained sequence, allowing us to engineer the three major reaction classes: non-catalytic, catalytic, and autocatalytic. These are the building blocks for composition of complex CRNs.

Extensive tests of the most basic reaction A + B → C verified correct stoichiometry (are the correct amounts of reactants used up and products generated? Fig. 2.3a) and kinetics (are the reactants and products being generated according to the target rate law?, Section Verification of the bimolecular rate law). In the catalytic reaction A + B → C + B, even a small amount of B effectively converts all of A to C, but B remains conserved (Fig. 2.3b). Catalytic reactions are ubiquitous in biological chemical controllers (e.g. transcriptional networks, kinase networks) as well as man-made artificial systems[21, 40, 37, 25, 41, 84]. We quantitatively analyse the catalytic turnover, showing that a single catalyst can trigger multiple reaction cycles (see Fig. 2.4).

In the autocatalytic reaction A + B → C + 2B even a small amount of B effectively converts all of A to itself (C acts as a read-out), resulting in the typical sigmoidal kinetic curves (Fig. 2.3c). Because of the exponential growth kinetics, autocatalytic reactions are common in settings where rapid (self-)amplification is observed such as replication or apoptosis. These properties also make autocatalysis a key ingredient to propagate information in proposed chemical algorithms[101] (see also Section Consensus algorithm). Because autocatalysis is extremely sensitive to leaks[15, 41, 84], it provides a good measurement of reactant quality. The estimated amount of autocatalyst B leaked (black trace, Fig. 2.3c,ii) is less than 2%, however this leak is exponentially amplified.

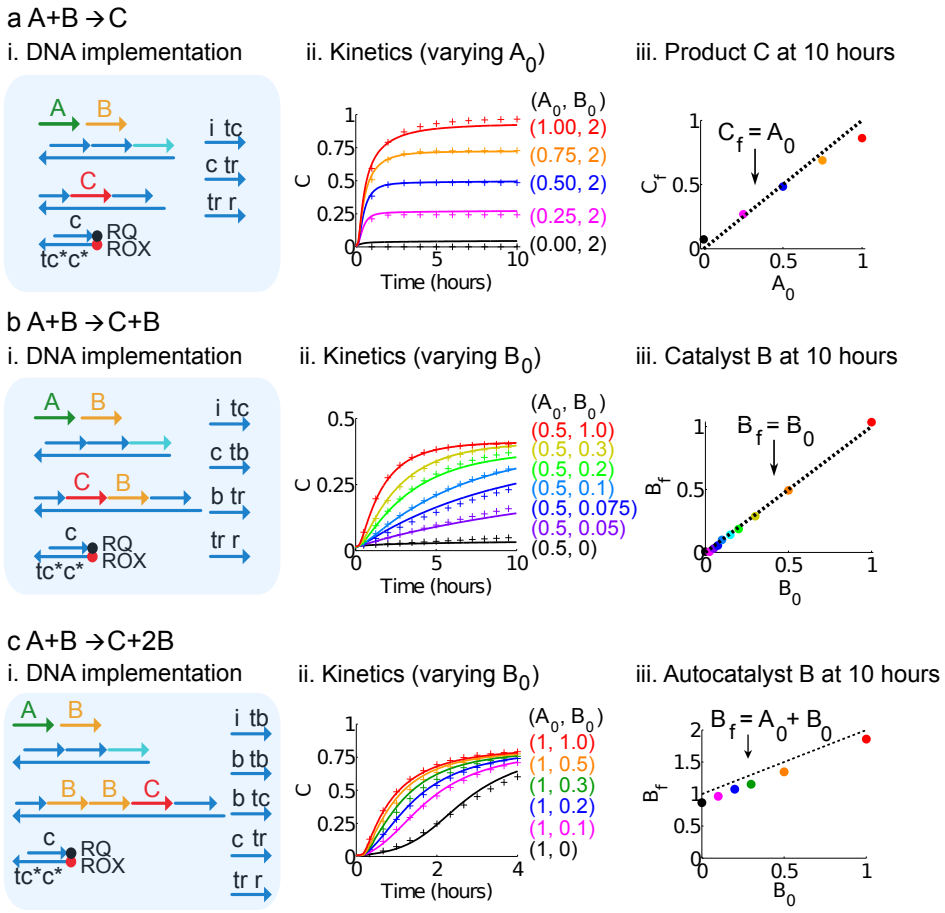


Figure 2.3: Testing fundamental reaction types. Panels (i) show a simplified representation of the gates, auxiliary strands, and signal strands used for the corresponding experiments. Experimental kinetics data are shown in panels (ii) as full coloured lines. Concentrations of the signal strands are indicated in the same colour, 1x = 50 nM. All join and fork gates were at 1.5x , and auxiliary strands were at 2x. Best fits of the strand displacement-level model to the data are shown as crossed lines. Panels (iii) show data confirming the correct reaction stoichiometry. **a,** Non-catalytic bimolecular reaction A + B → C. Signal B was at 2x and different amounts of signal strand A were added. Panel (iii) shows that levels of (product) signal C at the measurement end point (10 h) are very close to the amounts of limiting inputs as expected for a stoichiometrically correct bimolecular reaction. **b,** Bimolecular catalytic reaction A + B → C + B. Signal A was at 0.5x and different amounts of the catalytic signal B were introduced into the system. Panel (iii) shows that the final amount of free catalyst B_f is equal to the initial amount B_0. The amount of catalyst signal B at 10 h was measured by adding a fluorescent reporter for B. **c,** Autocatalytic reaction A + B → C + 2B. Signal A was at 1x and the amount of signal B was varied. Panel (iii) shows that the final amount of the autocatalyst signal B is equal to the sum of the initial amounts of A and B as expected for autocatalysis. The amount of autocatalyst signal B was measured at 10 h by adding a fluorescent reporter for B.

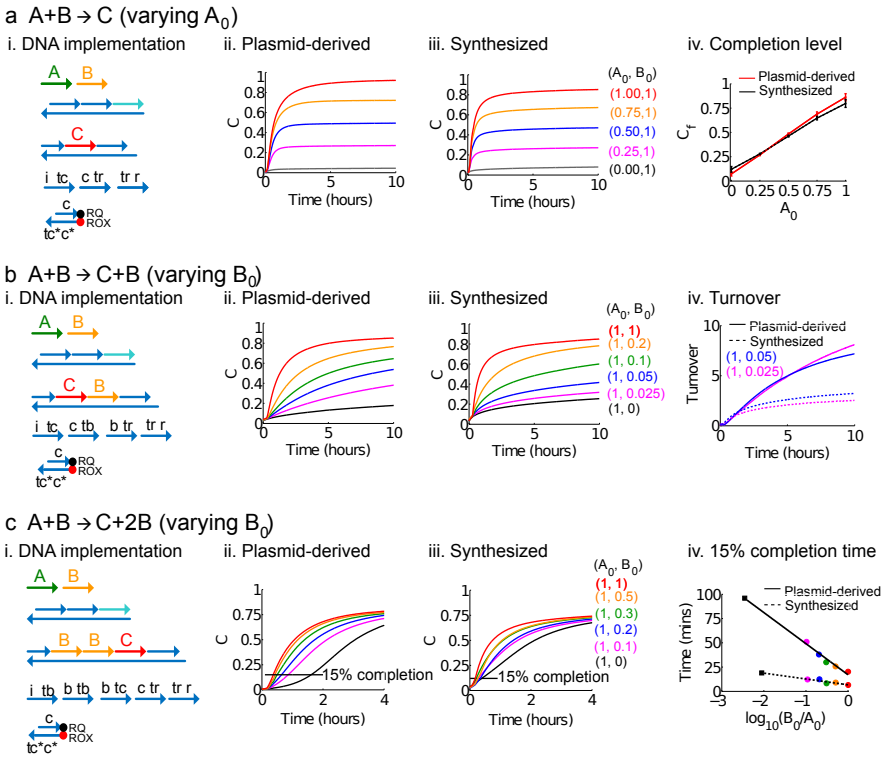


Figure 2.4: Kinetics data for plasmid-derived gates and data for synthesized gates. NdsDNA gates can be used to implement reactions with different kinetics. Panel (i) in each row shows a simplified representation of the ndsDNA gates and signal strands used for the corresponding experiments. Kinetics data for plasmid-derived gates are shown in panel (ii) and data for synthetic gates are shown in panel (iii). A fluorescent reporter for signal C was used to follow the reactions in all experiments. All join and fork gates were at 1.5x, 1x=50nM, and auxiliary strands were at 2x. Experiments were run in Tris-acetate-EDTA buffer containing 12.5mM Mg^{2+} (1x TAE/Mg^{2+}). **a,** Kinetics data for the bimolecular reaction A + B → C. Signal B was at 2x and different amounts of signal strand A were added as indicated in the figure. Plasmid-derived gates demonstrate similar kinetics to the gates assembled from synthetic DNA. The experiment was run at 25°C. Panel (iv) shows reaction completion levels. Endpoints for all data traces in the second and third panel are shown as a function of the concentration of signal A. **b,** Bimolecular catalytic reaction A + B → C + B. Signal A was at 1x and different amounts of the catalytic signal B were introduced into the system. The reaction was tested at 35°C. Panel (iii) shows that plasmid-derived gates exhibited higher turnover than synthesized gates. For this analysis, the leak reaction (0x input) was subtracted from the traces with 0.025x and 0.05x catalyst. Then, each trace was divided by the corresponding catalyst concentration to obtain a turnover number, that is a measure for the amount of product produced per unit of catalyst at a given time during the reaction. **c,** Autocatalytic reaction A + B → C + 2B. Signal A was at 1x and the amount of signal B was varied. Panel (iv) shows a linear fit of the 15% completion time against the logarithm of the relative concentration of signal B. The faster reaction speeds and thus higher slope observed for the synthesized system are likely due to the higher leak in that system.

While the use of plasmid-derived gates was found to improve circuit performance, an analysis of the cost and processing times revealed that while the production of plasmid-derived gates is slightly cheaper (Table 2.2), it takes 2-3 times longer processing time compared to assembly and purification of gates from commercially synthesized oligos (Table 2.1). The primary costs of plasmid-derived gates are gene synthesis and the use of restriction enzymes. For 300 pmole of gates (enough for 15 reactions at 30 nM), the estimated cost for Join gates is approximately $170 and $200 for Fork gates, the cost difference being due to the use of different nicking enzymes. In contrast, the chemical synthesis of the strands for the same gate costs around $260 including a PAGE purification fee. The primary time cost for plasmid-derived gates is in the cloning procedure, which, just like DNA synthesis, can be outsourced to a gene synthesis company. However, once assembled, plasmid-derived gates have the advantage that the host plasmids can easily be replicated and can be stored in the form of bacterial glycerol stocks. This makes it possible to reuse the gates many times over.

Table 2.1: Processing times comparison between plasmid-derived gates and synthesized gates

Plasmid-derived gates		**Synthesized gates**	
Processing	Time consumption	Processing	Time consumption
Cloning	∼5 hr	Annealing	∼1 hr
Plasmid extraction	∼2 hr	PAGE purification	∼2 hr
Enzyme digestion (two steps)	∼0.5 hr	Total	∼3 hr
Buffer purification (ethanol precipitation)	∼1 hr		
Total	∼8.5 hr		

Table 2.2: Cost comparison between plasmid-derived gates and synthesized gates.

Plasmid-derived gates		**Synthesized gates**	
Description	Cost	Description	Cost
Gate template	∼$100	PAGE ultramer (100 nt bottom strand)	∼$75
Plasmid extraction (QIAGEN)	∼$26	PAGE purified DNA oligo (30 nt top strand)	∼$185
Restriction enzyme	∼$11 (PvuII-HF)	Total	∼$260
Nicking enzyme	∼$62 (Nt.BstNBI, Fork gate)		
	∼$29 (Nb.BsrDI, Join gate)		
Total	∼$199 (Fork gate)		
	∼$166 (Join gate)		

2.6 Verification of the bimolecular rate law.

The reaction specification A + B → C delineates not only the production/consumption relationships between A, B, C, but also the dynamics. Despite the overall complex reaction mechanism (which, for A + B → C, involves 5 reversible and 1 irreversible stand displacement reactions, Fig. 2.5b), an analytical argument shows that the overall kinetics should be well approximated by the mass-action rate law expected of the formal reaction (i.e. d[C]/dt=-d[A]/dt=-d[B]/dt=k[A][B]). As the derivation (see section 4.1) shows, the regime of best correspondence (CRN regime) is one in which gates and auxiliary strands, including backward auxiliary strands <a tb> and <b tr>, are sufficiently in excess over the signal strands Fig. 2.5.

We experimentally confirmed that the multi-step strand displacement level mechanism implements the expected rate law for A + B → C, and that the rate constant can be tuned by adjusting the concentrations of gates and auxiliary species. Fig. 2.5b shows six sets of experimental data for the reaction A + B → C in or near the CRN regime. Each set was obtained with a different concentration of the backward auxiliary strands <a tb> and <b tr> and contains kinetics traces corresponding to at least two different combinations of the signals A and B. We chose to vary the concentration of the backward auxiliary strands because our analysis suggests that the formal rate constant can be effectively tuned in this way. We then fit the data from each set to a bimolecular rate law; the best fit rate constants varied over about two orders of magnitude from $3.5*10^4$/M/s to $5.3*10^2$/M/s as the concentration of backward auxiliary strands increased from 0x to 13x. The data show that the reactions are symmetrical with regard to the two signals as required by the bimolecular rate law although the signal strands A and B react sequentially with the join gate (see e.g. traces with A, B at 1x, 0.3x and 0.3x, 1x respectively).

2.7 Mechanistic strand displacement-level model.

Each individual strand displacement step can be well modelled as a bimolecular reaction between a signal or auxiliary strand and a gate complex with a matching open toehold[18]. We used the Visual DSD[85, 102] software to develop a quantitatively constrained model of the dynamics of our system on this mechanistic level. We allowed each strand displacement step to occur at a different rate depending on the sequences of the toeholds and adjacent domains. The model includes a phenomenological treatment of the erroneous leak reactions (Supplementary text S7.3 in [35]). We fit all the data that we obtained for the non-catalytic, catalytic and autocatalytic reactions (Figs. 2.3 and 2.5), and independent measurements of a large number of intermediate reaction steps (Figs. S16-S17 in [35]). These 104 data traces yielded a highly constrained set of strand displacement rate constants, with values ranging from 10^4 /M/s to $1.44*10^6$ /M/s (Table S3 in [35]), consistent with previously reported data[18]. The mechanistic strand displacement-level model fitted the data for all reaction conditions, including leak reactions, exceptionally well (crossed lines in Fig. 2.5 and Fig. 2.5b, see also SI Section S7 in [35]).

The strand displacement rate constants can be used to predict the effective bimolecular rate constant for the target formal reaction A + B $\rightarrow$ C (see section 4.1). We compared this predicted rate constant to that obtained by direct fitting of the data in Fig. 2.5b to a bimolecular rate law. Fig. 2.5c shows that our prediction is in good agreement as long as the concentration of the backward strands is 3x or higher. Divergence is expected at lower concentrations since the approximation we made to derive the analytic result does not hold. These results demonstrate that we can systematically vary the formal rate constant through quantitative control over the underlying reaction mechanism.

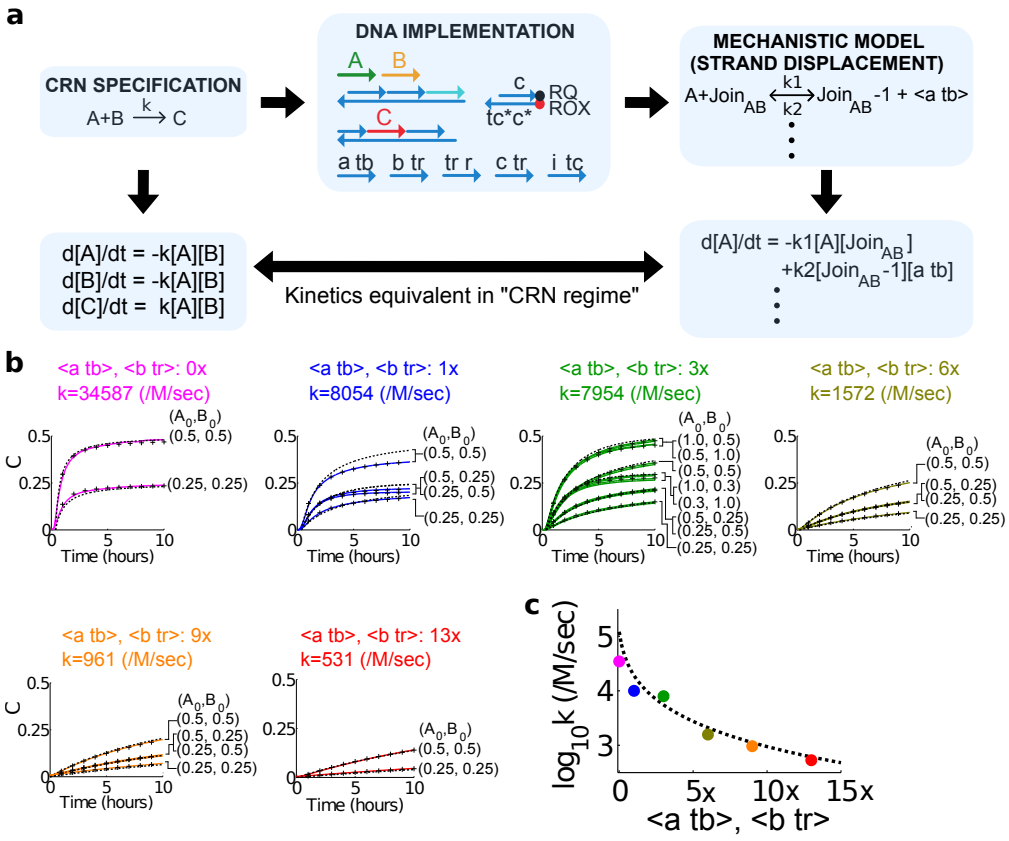


Figure 2.5: Tuning the rate of the bimolecular reaction A + B → C. a, Approximating the bimolecular rate law. The CRN program is executed by a DNA architecture that can be quantitatively modelled at a mechanistic level. We view strand displacement reactions (for example, A + Join_{AB} → Join_{AB}-1 + <a tb>, see Fig. 2.1b for component names) as the elementary reaction steps and the formal bimolecular reaction (for example, A + B → C) as the complex reaction pathway decomposed into these elementary reactions. In the CRN regime (see text) the mechanistic model closely approximates the dynamics of the target program. The rate constant k of the formal system can be tuned by changing the concentrations of gates and auxiliary strands. **b,** Reactions with varying concentrations of the backward auxiliary strands <a tb> and <b tr>. The data (solid traces) show the time evolution of C; purple traces (0x <a tb> and <b tr>), blue (1x), green (3x), olive green (6x), orange (9x) and red (13x), where 1x = 40 nM. Gates were at 3x and the initial concentrations of signals A and B are indicated in each panel. Black dashed lines are fits to the bimolecular rate law in a. Best-fit rate constants to the bimolecular rate law are indicated in each panel. Black crossed lines are fits to the mechanistic strand displacement-level model. **c,** Fitted bimolecular rate constant versus analytic prediction. The solid line is obtained from an analytic prediction for the dependence of the expected rate constant on the concentrations of the backward auxiliary strands <a tb> and <b tr>. The coloured dots show the rate constants k obtained from fitting the experimental data from **b**.

2.8 Consensus network.

An important function of molecular controllers is their ability to make decisions by comparing concentrations of input signals (Fig. 2.6a). We engineered such a decision-making controller by implementing a consensus algorithm that operates on two signals (X and Y). The signal that is initially in the minority is completely eliminated and replaced by the signal that is initially in the majority (in any given experiment the sum of all signal concentrations is constant) (Fig. 2.7). In distributed computing parlance, we implement an algorithm that allows picomole (10^{11}) quantities of agents, each with vote X or Y, to agree on a majority decision[101, 103]. The classification into minority and majority is thus unequivocal, distinguishing this network from previous proposals for DNA-based molecular classifiers[104] where the resulting signal was proportional to the difference in the initial concentrations (which can be small when concentrations are close). Consensus is a basic distributed computing problem and provides for us a proof-of-principle that CRN algorithms are directly translatable to our DNA controllers.

The network consists of two autocatalytic and one non-catalytic bimolecular reaction (see Fig. 2.6b): A reaction between signals X and Y creates two copies of the buffer signal B while a reaction involving B and X (Y) creates two copies of X (Y). Intuitively, the minority and majority signals initially cancel each other producing the buffer signal, which is then converted back to the majority signal. For the threshold to be set at the point where concentrations of the two species are equal, the two autocatalytic reactions should occur at the same rate. To compensate for minor differences in the rates of the two autocatalytic reactions (Fig. 2.8) and make reaction rates similar, we adjusted gate and auxiliary species concentrations.

Fig. 2.8 shows the ndsDNA gates needed to implement this reaction network. Example data traces in Fig. 2.6c clearly show that the network correctly classifies the majority for all eight input combinations tested. Each panel shows the time evolution of the concentration of X (red), Y (green) and B (yellow) for given starting concentrations of X and Y. Fig. 2.6d

shows the net amplification of the majority signal relative to the minority signal. The results for the DNA implementation and the expected dynamics of the formal CRN (see Fig. 2.7) are in qualitative agreement: we observe the gradual decrease of the minority, intermediate build-up of buffer and the initial decrease but long-term increase of the majority.

We also constructed a strand displacement model for each reaction of the consensus network using Visual DSD, and parameterized these models using experimental measurements for each reaction and for the individual fork and join gates (Figs. S21, S23, S24 in [35]). By composing models of individual reactions into a model of the full consensus network, we were able to quantitatively predict the dynamics of the consensus network solely from the models of its constituent parts, up to a constant scaling factor (Fig. 2.6c, see Section S8 in [35] for further details).

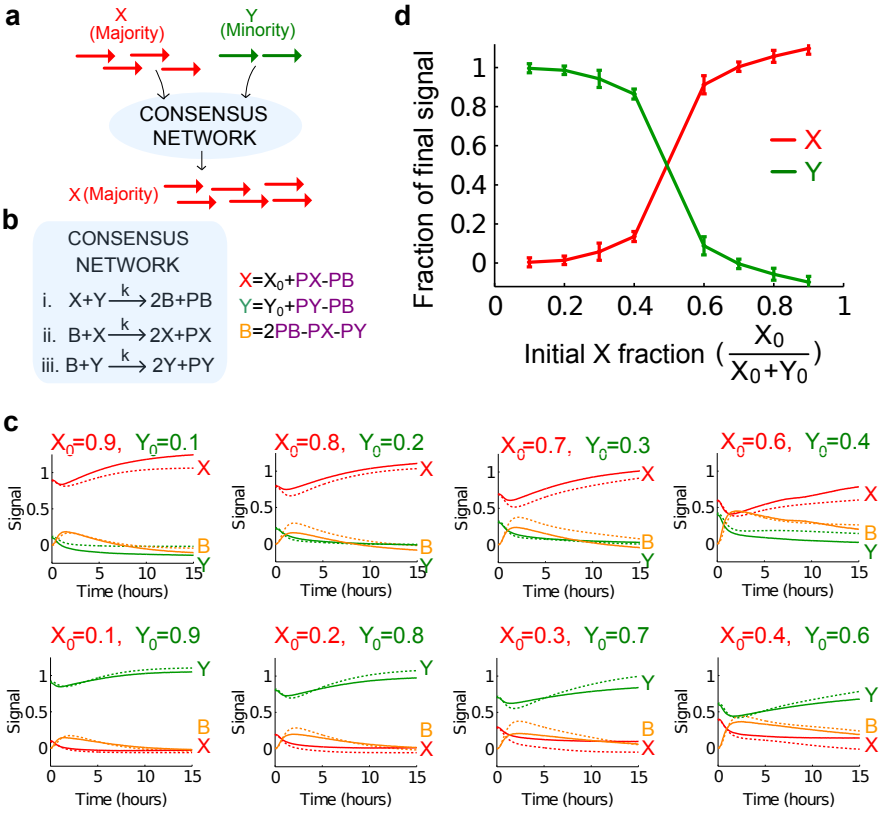


Figure 2.6: Consensus network. a, Given arbitrary amounts of signal strands X (red) and Y (green), the consensus network converts the minority signal to the majority signal. **b,** The formal chemical reactions for the consensus network. Signals PX, PY and PB were used to follow the reaction kinetics without interfering with the dynamics of X, Y and B. Reporters for PX, PY and PB each used a different fluorophore such that all three signals could be detected in the same reaction. The values of X, Y and B were calculated from the measured values of PX, PY and PB as indicated. **c,** Time evolution of signals X (red), Y (green) and B (yellow). Initial concentrations of signals X and Y are indicated in each panel, 1x = 80 nM. Reporters were at 3x , auxiliary strands at 2x and gates at 2x for reactions (i) and (ii). Gates and auxiliary strands for reaction (iii), B + Y right arrow 2Y, were at 2.4x to balance the rates of the two autocatalytic reactions. The DNA implementation for the consensus network consisted of 3 join gates, 3 fork gates, 3 reporters, 13 auxiliary strands and 3 signal strands. No backward auxiliary strands were added to the initial reaction mixture. The kinetics data show that the minority species was converted into the buffer species B first, then into the majority species. The model prediction of the consensus network using the strand displacement-level model is shown as dashed lines. The prediction is based on a model parameterization obtained by fitting to the individual reactions. **d,** Amplification levels. The end points (15 h) of each reaction show that the DNA-based consensus network correctly amplifies the majority towards totality. Red trace: X/(X + Y) at 15 h; green trace: Y/(X + Y) at 15 h. Error bars indicate standard deviation calculated from three independent experiments.

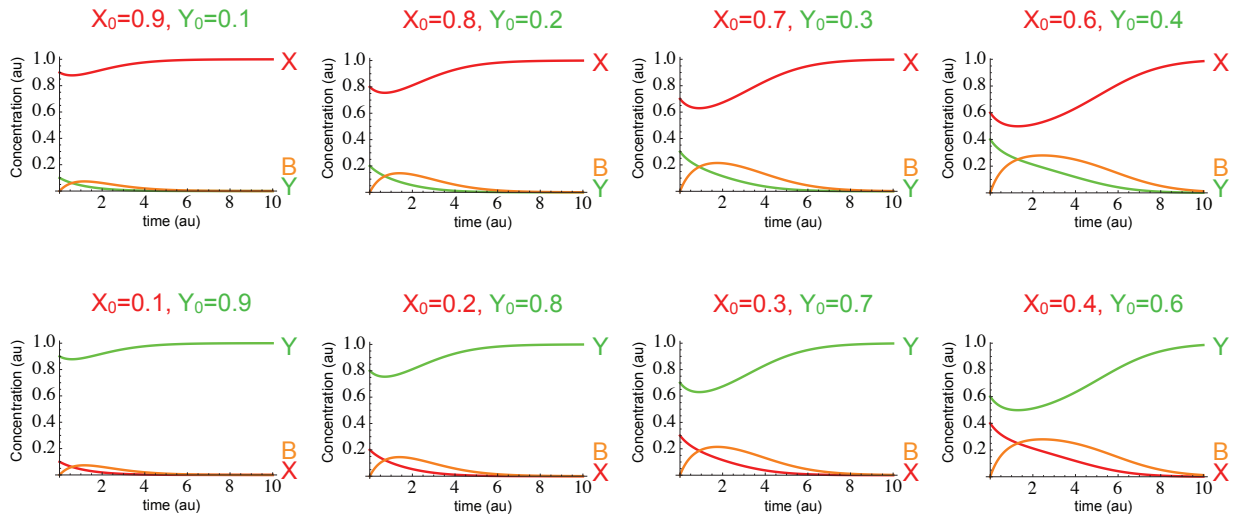


Figure 2.7: Ideal consensus network behavior. Unitless numerical simulation of the consensus network in Fig. 2.6b, using k=1. The behavior of the system qualitatively matches experimentally obtained data shown in Fig. 2.6: both the majority and minority species initially drop to form B and then the majority species recovers while the minority continues to decay.

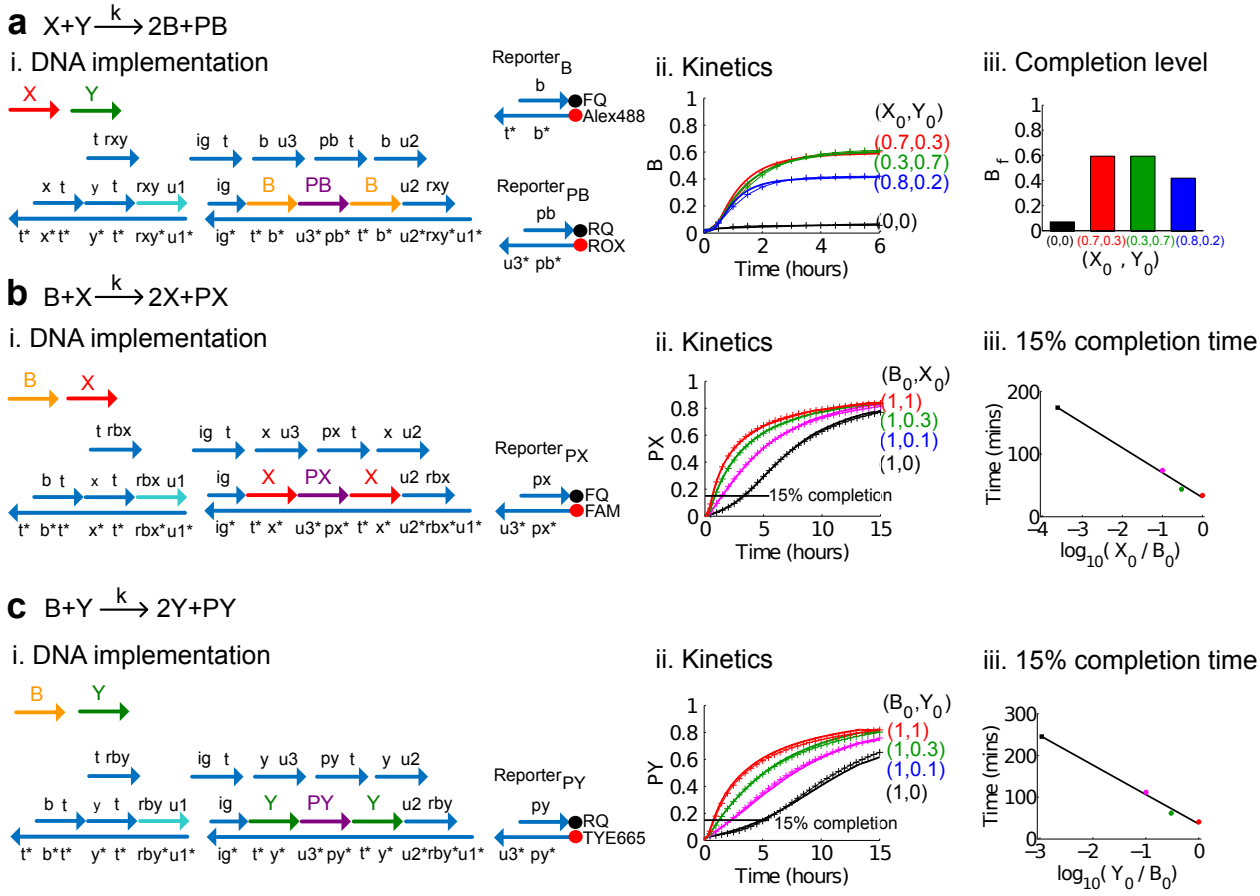


Figure 2.8: Kinetics data for individual chemical reactions for the consensus network. Panel i in each row shows the signal strands, auxiliary strands, ndsDNA gates and reporter gates used for the corresponding experiments. Experiments were run in 1x TAE/Mg^{++} and performed at 25°C. All auxiliary strands were at 100 nM (2x), and the gates were at 75 nM concentration (1.5x). Best fits of the strand displacement-level model to the data are shown as crossed lines. **a**, X+Y→2B+PB. Different amounts of signal X and signal Y were added. The $\text{Reporter}_{\text{B}}$ was used for the characterization of X+Y→2B+PB, and the $\text{Reporter}_{\text{PB}}$ was used for the consensus network. Panel ii shows the kinetics of the signal B. Panel iii shows the completion level of signal B. **b**, B+X→2X+PX. Signal B was fixed at 50 nM (1x). Different amounts of signal X were added. Panel ii shows the kinetics of the signal PX. Panel iii shows a linear fit of the 15% completion time against the logarithm of the relative concentration of signal PX. **c**, B+Y→2Y+PY. Signal *B* was fixed at 50 nM (1x). Different amounts of signal Y were added. Panel ii shows the kinetics of the output PY. Panel iii shows a linear fit of the 15% completion time against the logarithm of the relative concentration of signal PY.

2.9 Conclusions.

We have developed a new systematic design strategy for non-living molecular systems with functional behaviours, paired with a technology for robust and effcient synthesis of the molecular components. Our scheme is built upon *de novo* designed interactions not known to occur in nature. Our components did not require (directed) evolution to achieve efficacy, but were designed in their ultimate form by the authors. As such, our work can be seen as a step in the larger human enterprise of recapitulating the mastery of biology over matter with *de novo* engineering.

As human engineering is driven inexorably towards molecular-scale devices, we must be careful to avoid shoehorning theory developed for digital electronics (for example, logic circuits) into the chemical context. For well over a century, CRNs have provided the mathematical language to describe and predict the dynamics of chemical experiments. Here, we leveraged this significant theoretical investment and demonstrated the prescriptive use of CRNs for programming molecular nanocontrollers.

Although our devices are entirely synthetic, they are biocompatible and there is a natural path towards applications in sensing and smart drug-delivery *in vivo*. Cell state is encoded in the sequences and concentrations of RNAs, and recent work has shown that strand displacement logic gates can recognize miRNA profiles in living mammalian cells [79]. Further, the demonstration of the power of strand displacement as a mechanism for building synthetic molecular circuits tempts the hypothesis that there are natural strand displacement-based cellular regulatory networks with interesting dynamics yet to be discovered.

2.10 Methods.

2.10.1 Preparation of plasmid-derived ndsDNA gates.

Double-stranded DNA templates were cloned into a high-copy-number plasmid and transformed into E. Coli. A single colony was picked from an Ampicillin selective plate and a 800 ml overnight culture was grown with Ampicillin at 100 μg/ml at 37°C with vigorous shaking. Plasmids were extracted using a QIAGEN Maxi-prep kit and inserts were sequenced to ensure that there was no sequence error or recombination in the ndsDNA gates. Cloned ndsDNA gates were first digested with a restriction enzyme (PvuII-HF) at 37°C for one hour to release the gates from the plasmid backbone. Reactions were run with 4 units of PvuII-HF per 1 μg of plasmid. The reaction mix then was ethanol precipitated in order to optimize reaction conditions for the next digestion step. Join gates were digested with the nicking enzyme Nb.BsrDI at 65°C for one hour using 4 units of enzyme per 1 μg of plasmid. Fork gates were digested with the nicking enzyme Nt.BstNBI at 55°C for one hour using 8 units of enzyme per 1 μg of plasmid. All enzymes were purchased from New England Biolabs. For kinetics experiments, enzymes were dissociated from DNA by adding sodium dodecyl sulfate (SDS) to a final concentration of 0.15% (see Fig. 2.9). Gates were then used for experiments without further purification from enzymes or plasmid backbone.

2.10.2 Modeling and parameter inference.

Computational models were constructed for each analyzed circuit, using the DNA Strand Displacement (DSD) programming language and Visual DSD software ([85, 102]). The unknown kinetic parameters in the model were inferred from the experimental data using Markov Chain Monte Carlo methods, as implemented in the Filzbach software see authors website: http://research.microsoft.com/science/tools. Such methods require the definition of a likelihood function, which describes the probability of reproducing the observed data D, given a model hypothesis H and corresponding parameter set θ, i.e. $\Pr(D|\theta,H)$ (see [105] for a comprehensive description of Monte Carlo statistical methods). We used or-

dinary differential equation simulations for each circuit. C# code was generated using the Visual DSD tool, then integrated numerically with adaptive step-size ODE integrators (http://mstlab.org/eng/projects/Pages/Solvers.aspx Microsoft Research Solvers library for .NET).

2.10.3 Sequence design.

Strand displacement circuitry using double-stranded DNA was proposed in Ref. [3] by Cardelli. Here, we provide the first experimental verification of this design. Ref. [3] proposed a design where all toeholds have the same sequence but in our experimental work we used a variant of the original design where different toeholds have different sequence. The nicking enzyme recognition sites do not coincide with the toeholds but are part of the long domain sequences. (Nb.BsrDI and Nt.BstNBI recognition sites are on the left and right side of the long domain, respectively.) The use of a nicking enzyme to make breaks in one of the gate strands imposes certain sequence constraints but these constrains can be incorporated without loss of generality. There are enough unconstrained nucleotides in each strand or domain to ensure that all sequences can be made orthogonal, minimizing crosstalk. For a very large reaction network it may be desirable to further increase these differences. In this case, we can simply extend the length of the long domains and insert additional unconstrained nucleotides .

2.10.4 DNA synthesis and purification

DNA oligos were purchased from Integrated DNA Technologies (IDT). Inserts for gate cloning were ordered as double-stranded gBlocks. Long bottom strands of synthesized ndsDNA gates were ordered as polyacrylamide gel electrophoresis (PAGE) purified ultramers. Signal species X and Y were ordered purified by high-performance liquid chromatography (HPLC), and reporter strands with fluorophore/quencher modifications were also HPLC purified. Other single-stranded DNA were ordered as PAGE purified. Strands were re-suspended and 100 μM stock solutions were prepared in 1x TE.

2.10.5 Preparation of synthesized gates

Join and fork gates were prepared from stock DNA solutions with 20% excess of top strands, in 1x Tris-acetate- EDTA buffer containing 12.5mM Mg^{++} (1x TAE/Mg^{++}). Resulting gate concentrations were typically in the range of 20 μM and total amounts were approximately 3 nmole. Gates were annealed in a thermal cycler, cooling from 95°C to 20°C at a rate of 0.1°C /min. This slow annealing process ensured the correct formation of complexes. Reporter complexes were annealed with a 20% excess of top strands, in 1x TAE/Mg^{++}, and a faster anneal was used, going from 90°C to 20°C at a rate of 1°C /min in order to save time. The excess of top strands with quenchers ensured complete quenching of fluorescent bottom strands. Join and fork gates were further purified to remove excess single-stranded DNA and poorly formed gate complexes using non-denaturing (ND) PAGE (gel preparation see 2.10.8). The proper bands were cut from the gel and soaked in 1x TAE/ Mg^{++} for two days at room temperature to elute the DNA.

2.10.6 Kinetics experiments and fluorescence data normalization

Kinetics experiments were performed on a spectrofluorimeter (SPEX Fluorolog-3, Horiba) with 0.875 ml synthetic quartz cells (Starna catalog number 23-5.45-S0G-5). We used a 2.73 nm slit width for both excitation and emission monochromators. Excitation and emission wavelengths of different fluorophores were as follows: ROX (588 nm/ 608 nm), FAM (495 nm/ 520 nm), Cy3 (550 nm/ 564 nm), TYE665 (645 nm/ 665 nm), TAMRA (559 nm/ 583 nm), AL488 (492 nm/ 517 nm), and AL647 (650 nm/ 670 nm). The integration time was 10 sec for all experiments for every 60 sec time-point. Since DNA has been observed to non-specifically bind to pipette tips, a non-reactive 20 nt poly-T carrier strand at the concentration of 1 uM was used in all reactions so that the majority of DNA loss would be the carrier strands. Addition of carrier strands could improve the measurement accuracy. In all kinetics experiments, arbitrary fluorescence units were converted to the concentrations of the corresponding signal strand using a calibration curve of each reporter complex. Calibration

experiments were run once for each reporter complex.

2.10.7 Gate concentration quantification

The ndsDNA gate complex can be fully triggered by adding excess of all necessary signal and auxiliary strands, and then the concentration of the gate complex can be determined by reading off the produced fluoroscence signal. For example, in the case of Fork_C, a small amount of Fork_C ($\sim$0.5x) was mixed with 3x of the Reporter_C. Then excess amounts ($\sim$10x) of all necessary auxiliary strands (i.e. <i tc>, <c tr>, <r tq>) were added to trigger the Fork_C gate, After the reaction reached equilibrium, the concentration of Fork_C can be determined by reading off the concentration of the produced signal C.

2.10.8 Gel electrophoresis

Non-denaturing PAGE

10% non-denaturing PAGE gels were made by mixing 10 ml 19:1 40% acrylamide/bis, 4 mL 10x TAE/Mg^{2+}, and deionized water to 40 mL. Then 300 μL APS and 30 μL TEMED were added to help polymerization. 80% glycerol was added to all annealed samples, achieving final glycerol concentration of 15% by volume. The samples were run on 10% ND PAGE at 140 V for 6 hrs at 25°C.

Denaturing PAGE

10% denaturing PAGE gels were made by mixing 10 mL 19:1 40% acrylamide/bis, 4 mL 10x TBE, 16.8g urea, and deionized water to 40 mL, then adding 300 μL APS and 30 μL TEMED were added to help polymerization. For plasmid-derived sample, we loaded approximately 1 μg of DNA (including gates and plasmid backbone) and for synthetic DNA we loaded 150-200 ng of gate complex. 2x TBE/Urea denaturing loading buffer(Bio-Rad) was added to all samples in 1:1 ratio. Gels were run at 120 V for 3 hours at 55°C with the temperature controlled using an external temperature bath. Gels were stained with Sybr-Gold stain

(Invitrogen) at room temperature for 20 minutes, and scanned with a Typhoon FLA 9000 biomolecular imager.

2.10.9 Cloning stategy of plasmid-derived ndsDNA gates

In order to increase the yield of plasmid-derived ndsDNA gates, we cloned multiple identical ndsDNA gates into one plasmid. ndsDNA gate templetes were ordered as doubles-stranded gBlocks from Integrated DNA Technologies (IDT). Four different pairs of primers were used to amplify a gate complex, and an extra 40 bp of overlapping sequence (red, orange, green, purple, and magenta) were added at each terminus (Fig. 2.10). The amplified parts were further purifed from 2% agarose gel. Purified fragments and the linearized vector were then combined with the Gibson Master Mix containing T5 exonuclease, Phusion DNA polymerase, and Taq ligase [106]. The reaction was incubated at 50°C for one hour, and during incubation, multiple copies of the gate complexes were assembled into the plasmid backbone.

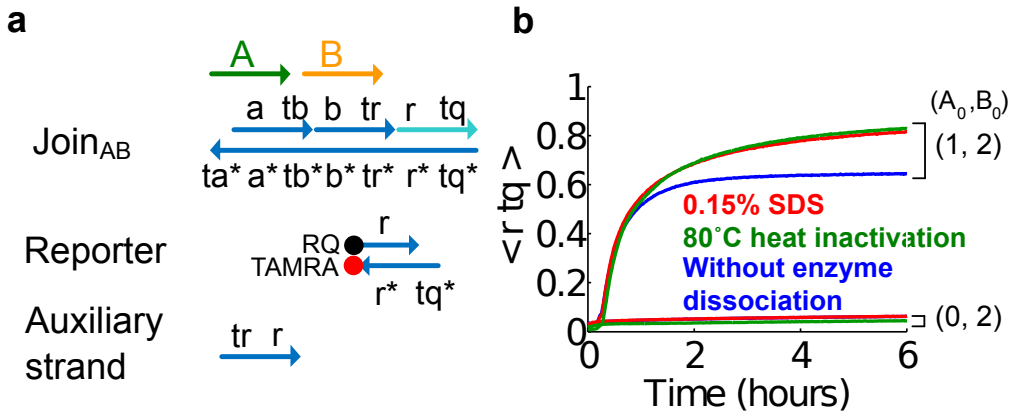


Figure 2.9: Enzyme dissociation and circuit behavior. a, Diagram of the ndsDNA gate tested here. **b,** Kinetics experiments of the plasmid-derived Join$_{AB}$ using 80°C heat inactivation (green traces), 0.15% sodium dodecyl sulfate (SDS) (red), and a control without heat inactivation or addition of SDS (blue). The standard concentration was 1x=10nM, and all auxiliary strands and input B were at 2x. The gate complex was 1.5x, and the experiments were performed at 35°C in Tris-acetate-EDTA buffer containing 12.5 mM Mg^{2+}.

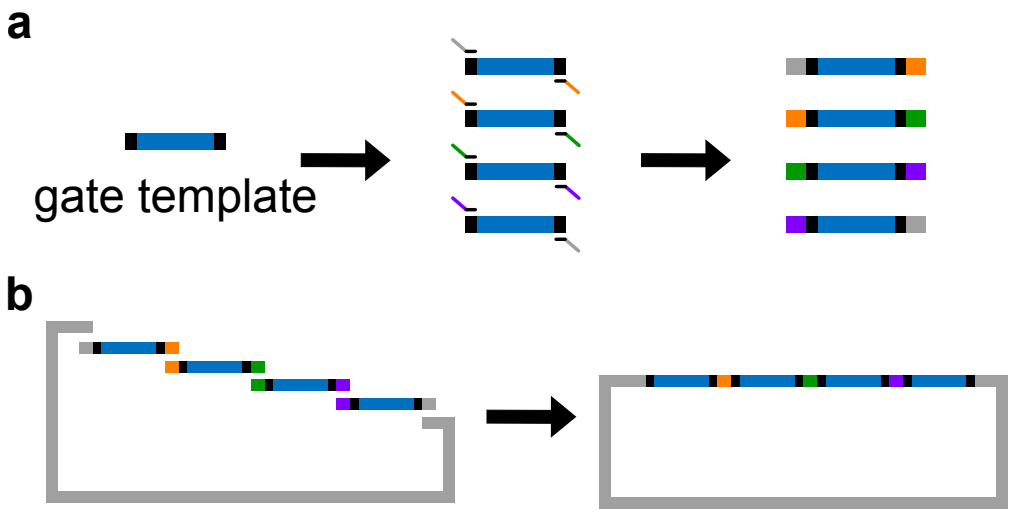


Figure 2.10: Cloning strategy of plasmid-derived ndsDNA gates. a, PCR of a DNA gate template. A DNA gate template contains the ndsDNA gate sequences in the center (a blue region), and spacer sequences on both ends (black regions; these two end sequences are orthogonal). Primers can bind to the spacer sequences of the gate template, and generate four overlapping DNA fragments through PCR (overlapping sequences are color-coded in the figure). **b,** Gibson assembly. The four amplified DNA fragments are then assembled into a linearized plasmid backbone through Gibson assembly method[106].

2.11 Acknowledgements.

The authors thank E. Winfree, E. Klavins and D.Y. Zhang for discussions and comments on the manuscript. This work was supported by the National Science Foundation (grant NSF-CCF 1117143 to G.S. and D.S.). G.S. was supported by a Burroughs Wellcome Career Award at the Scientific Interface. D.S. was supported by an NIGMS Systems Biology Center grant (P50 GM081879).

Table 2.3: Domain sequences of ndsDNA gates used in Fig. 2.3. All sequences start from the 5′ end.

Domain	**Sequence**	**Length (nt)**
ta	CTGCTA	6
tb	TTCCAC	6
tc	TACCCA	6
tr	TCCTAC	6
tq	AACCAG	6
a	CATTGCTTCTACGAGTCATCC	21
b	CATTGCACCTTAGAGTCCGAA	21
c	CATTGCCACATCGAGTCCCTT	21
r	CATTGCTTAACCGAGTCTCAC	21
i	CTGCCATCATAAGAGTCACCA	21

Table 2.4: Strand sequences of ndsDNA gates used in Fig. 2.3. Domain sequences are in Table. 2.3. All sequences start from the 5′ end.

Strand	**Domain**	**Sequence**	**Length (nt)**
Join_{AB}- Bottom	tq* r* tr* b* tb* a* ta*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA TTCGGACTCTAAGGTGCAATG GTGGAA GGATGACTCGTAGAAGCAATG TAGCAG	87
Fork_{C}- Bottom	tq* r* tr* c* tc* i*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA AAGGGACTCGATGTGGCAATG TGGGTA TGGTGACTCTTATGATGGCAG	81
Fork_{BC}- Bottom	tq* r* tr* b* tb* c* tc* i*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA TTCGGACTCTAAGGTGCAATG GTGGAA AAGGGACTCGATGTGGCAATG TGGGTA TGGTGACTCTTATGATGGCAG	108
Fork_{CBB}- Bottom	tq* r* tr* c* tc* b* tb* b* tb* i*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA AAGGGACTCGATGTGGCAATG TGGGTA TTCGGACTCTAAGGTGCAATG GTGGAA TTCGGACTCTAAGGTGCAATG GTGGAA TGGTGACTCTTATGATGGCAG	135
Fork_{BCB}- Bottom	tq* r* tr* b* tb* c* tc* b* tb* i*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA TTCGGACTCTAAGGTGCAATG GTGGAA AAGGGACTCGATGTGGCAATG TGGGTA TTCGGACTCTAAGGTGCAATG GTGGAA TGGTGACTCTTATGATGGCAG	135

Fork$_{BBC}$-Bottom	tq* r* tr* b* tb* b* tb* c* tc* i*	CTGGTT GTGAGACTCGGTTAAGCAATG GTAGGA TTCGGACTCTAAGGTGCAATG GTGGAA TTCGGACTCTAAGGTGCAATG GTGGAA AAGGGACTCGATGTGGCAATG TGGGTA TGGTGACTCTTATGATGGCAG	135
<ta a>	ta a	CTGCTA CATTGCTTCTACGAGTCATCC	27
<tb b>	tb b	TTCCAC CATTGCACCTTAGAGTCCGAA	27
<tc c>	tc c	TACCCA CATTGCCACATCGAGTCCCTT	27
<a tb>	a tb	CATTGCTTCTACGAGTCATCC TTCCAC	27
<b tr>	b tr	CATTGCACCTTAGAGTCCGAA TCCTAC	27
<r tq>	r tq	CATTGCTTAACCGAGTCTCAC AACCAG	27
<i>	i	CTGCCATCATAAGAGTCACCA	21
<tr r>	tr r	TCCTAC CATTGCTTAACCGAGTCTCAC	27
<i tc>	i tc	CTGCCATCATAAGAGTCACCA TACCCA	27
<c tr>	c tr	CATTGCCACATCGAGTCCCTT TCCTAC	27
<c tb>	c tb	CATTGCCACATCGAGTCCCTT TTCCAC	27
<b tr>	b tr	CATTGCACCTTAGAGTCCGAA TCCTAC	27
<i tb>	i tb	CTGCCATCATAAGAGTCACCA TTCCAC	27
<b tb>	b tb	CATTGCACCTTAGAGTCCGAA TTCCAC	27
<b tc>	b tc	CATTGCACCTTAGAGTCCGAA TACCCA	27
<c tr>	c tr	CATTGCCACATCGAGTCCCTT TCCTAC	27
<b tr>	b tr	CATTGCACCTTAGAGTCCGAA TCCTAC	27
TYE665-<a* ta*>	a* ta*	/5TYE665/ GGATGACTCGTAGAAGCAATG TAGCAG	27
<a>-RQ	a	CATTGCTTCTACGAGTCATCC /3IAbRQSp/	21

Cy3-<b* tb*>	b* tb*	/5Cy3/ TTCGGACTCTAAGGTGCAATG GTG-GAA	27
<b>-FQ	b	CATTGCACCTTAGAGTCCGAA /3IABkFQ/	21
ROX-<c* tc*>	c* tc*	/56-ROXN/ AAGGGACTCGATGTGGCAATG TGGGTA	27
<c>-RQ	c	CATTGCCACATCGAGTCCCTT /3IAbRQSp/	21
<tq* r*>	tq* r*	CTGGTT GTGAGACTCGGTTAAGCAATG /36-TAMTSp/	27
RQ-<r>	r	/5IAbRQ/ CATTGCTTAACCGAGTCTCAC	21

Table 2.5: Strands occur in ndsDNA gates used in Fig. 2.3. Strand sequences see Table. 2.4.

Gate	**Strand**	**Bottom strand length (nt)**
Join_{AB}	Join_{AB}-Bottom, <a tb>, <b tr>, <r tq>	87
Fork_{C}	Fork_{C}-Bottom, <i>, <tc c>, <tr r>	81
Fork_{BC}	Fork_{BC}-Bottom, <i>, <tc c>, <tb b>, <tr r>	108
Fork_{CBB}	Fork_{CBB}-Bottom, <i>, <tb b>, <tb b>, <tc c>, <tr r>	135
Fork_{BCB}	Fork_{BCB}-Bottom, <i>, <tb b>, <tc c>, <tb b>, <tr r>	135
Fork_{BBC}	Fork_{BBC}-Bottom, <i>, <tc c>, <tb b>, <tb b>, <tr r>	135
Reporter-<ta a>	TYE665-<a* ta*>, <a>-RQ	27
Reporter-<tb b>	Cy3-<b* tb*>, <b>-FQ	27
Reporter-<tc c>	ROX-<c* tc*>, <c>-RQ	27
Reporter-<r tq>	<tq* r*>-TAMRA, RQ-<r>	27

Table 2.6: Domain sequences of ndsDNA gates for the consensus network. All sequences start from the 5′ end.

Domain	**Sequence**	**Length (nt)**
t	CTGATC	6
u1	CTTCAG	6
u2	CCATAC	6
u3	ATACCC	6
x	CATTGCTTTATTTACCGAGTCTTAT	25
y	CATTGCCTAACCCACCGAGTCCTTT	25
b	CATTGCCAATTCCTACGAGTCTACC	25
pb	CATTGCATTATATTCCGAGTCCTAC	25
px	CATTGCCTTCCCACTAGAGTCTCAC	25
py	CATTGCACCACCCTAAGAGTCTAAC	25
rxy	CATTGCTACCACCTCCGAGTCTAAC	25
rbx	CATTGCCAAACCATTAGAGTCAAAC	25
rby	CATTGCACCCTAATACGAGTCTCAC	25
ig	CTGAAATAAATAAATAGAGTCTACC	25

Table 2.7: Strand sequences of ndsDNA gates for the consensus network. Domain sequences are in Table. 2.6. All sequences start from the 5′ end.

Strand	**Domain**	**Sequence**	**Length (nt)**
Join_{XY}-Bottom	u1* rxy* t* y* t* x* t*	CTGAAG GTTAGACTCGGAGGTGGTAGCAATG GATCAG AAAGGACTCGGTGGGTTAGGCAATG GATCAG ATAAGACTCGGTAAATAAAGCAATG GATCAG	99
Join_{BX}-Bottom	u1* rbx* t* x* t* b* t*	CTGAAG GTTTGACTCTAATGGTTTGGCAATG GATCAG ATAAGACTCGGTAAATAAAGCAATG GATCAG GGTAGACTCGTAGGAATTGGCAATG GATCAG	99
Join_{BY}-Bottom	u1* rby* t* y* t* b* t*	CTGAAG GTGAGACTCGTATTAGGGTGCAATG GATCAG AAAGGACTCGGTGGGTTAGGCAATG GATCAG GGTAGACTCGTAGGAATTGGCAATG GATCAG	99
Fork_{BB}-Bottom	u1* rxy* u2* b* t* pb* u3* b* t* ig*	CTGAAG GTTAGACTCGGAGGTGGTAGCAATG GTATGG GGTAGACTCGTAGGAATTGGCAATG GATCAG GTAGGACTCGGAATATAATGCAATG GGGTAT GGTAGACTCGTAGGAATTGGCAATG GATCAG GGTAGACTCTATTTATTTATTTCAG	155
Fork_{XX}-Bottom	u1* rbx* u2* x* t* px* u3* x* t* ig*	CTGAAG GTTTGACTCTAATGGTTTGGCAATG GTATGG ATAAGACTCGGTAAATAAAGCAATG GATCAG GTGAGACTCTAGTGGGAAGGCAATG GGGTAT ATAAGACTCGGTAAATAAAGCAATG GATCAG GGTAGACTCTATTTATTTATTTCAG	155

Fork$_{YY}$-Bottom	u1* rby* u2* y* t* py* u3* y* t* ig*	CTGAAG GTGAGACTCGTATTAGGGTGCAATG GTATGG AAAGGACTCGGTGGGTTAGGCAATG GATCAG GTTAGACTCTTAGGGTGGTGCAATG GGGTAT AAAGGACTCGGTGGGTTAGGCAATG GATCAG GGTAGACTCTATTTATTTATTTCAG	155
<t x>	t x	CTGATC CATTGCTTTATTTACCGAGTCTTAT	31
<t y>	t y	CTGATC CATTGCCTAACCCACCGAGTCCTTT	31
<t b>	t b	CTGATC CATTGCCAATTCCTACGAGTCTACC	31
<u3 px>	u3 px	ATACCC CATTGCCTTCCCACTAGAGTCTCAC	31
<u3 py>	u3 py	ATACCC CATTGCACCACCCTAAGAGTCTAAC	31
<u3 pb>	u3 pb	ATACCC CATTGCATTATATTCCGAGTCCTAC	31
<x t>	x t	CATTGCTTTATTTACCGAGTCTTAT CTGATC	31
<y t>	y t	CATTGCCTAACCCACCGAGTCCTTT CTGATC	31
<b t>	b t	CATTGCCAATTCCTACGAGTCTACC CTGATC	31
<rxy u1>	rxy u1	CATTGCTACCACCTCCGAGTCTAAC CTTCAG	31
<rbx u1>	rbx u1	CATTGCCAAACCATTAGAGTCAAAC CTTCAG	31
<rby u1>	rby u1	CATTGCACCCTAATACGAGTCTCAC CTTCAG	31
<ig>	ig	CTGAAATAAATAAATAGAGTCTACC	25
<u2 rxy>	u2 rxy	CCATAC CATTGCTACCACCTCCGAGTCTAAC	31
<u2 rbx>	u2 rbx	CCATAC CATTGCCAAACCATTAGAGTCAAAC	31

<u2 rby>	u2 rby	CCATAC CATTGCACCCTAATACGAGTCTCAC	31
<ig t>	ig t	CTGAAATAAATAAATAGAGTCTACC CTGATC	31
<b u3>	b u3	CATTGCCAATTCCTACGAGTCTACC ATACCC	31
<pb t>	pb t	CATTGCATTATATTCCGAGTCCTAC CTGATC	31
<b u2>	b u2	CATTGCCAATTCCTACGAGTCTACC CCATAC	31
<x u3>	x u3	CATTGCTTTATTTACCGAGTCTTAT ATACCC	31
<px t>	px t	CATTGCCTTCCCACTAGAGTCTCAC CTGATC	31
<x u2>	x u2	CATTGCTTTATTTACCGAGTCTTAT CCATAC	31
<y u3>	y u3	CATTGCCTAACCCACCGAGTCCTTT ATACCC	31
<py t>	py t	CATTGCACCACCCTAAGAGTCTAAC CTGATC	31
<y u2>	y u2	CATTGCCTAACCCACCGAGTCCTTT CCATAC	31
FAM-<px* u3*>	px* u3*	/56-FAM/ GTGAGACTCTAGTGGGAAGGCAATG GGGTAT	25
<px>-FQ	px	CATTGCCTTCCCACTAGAGTCTCAC /3IABkFQ/	25
TYE665-<py* u3*>	py* u3*	/5TYE665/ GTTAGACTCTTAGGGTGGTGCAATG GGGTAT	31
<py>-RQ	py	CATTGCACCACCCTAAGAGTCTAAC /3IAbRQSp/	25
ROX-<pb* u3*>	pb* u3*	/56-ROXN/ GTAGGACTCGGAATATAATGCAATG GGGTAT	31

<pb>-RQ	pb	CATTGCATTATATTCCGAGTCCTAC /3IAbRQSp/	25
ALEX488-<b* t*>	b* t*	/5Alex488N/ GGTAGACTCGTAGGAATTGGCAATG GATCAG	31
<b>-FQ	b	CATTGCCAATTCCTACGAGTCTACC /3IABkFQ/	25
ROX-<x* t*>	x* t*	/56-ROXN/ ATAAGACTCGGTAAATAAAGCAATG GATCAG	31
<x>-RQ	x	CATTGCTTTATTTACCGAGTCTTAT /3IAbRQSp/	25
ALEX647-<y* t*>	y* t*	/5Alex647N/ AAAGGACTCGGTGGGTTAGGCAATG GATCAG	31
<y>-RQ	y	CATTGCCTAACCCACCGAGTCCTTT /3IAbRQSp/	25
<u1* rxy*>-TAMRA	u1* rxy*	CTGAAG GTTAGACTCGGAGGTGGTAGCAATG /36-TAMTSp/	31
RQ-<rxy>	rxy	/5IAbRQ/ CATTGCTACCACCTCCGAGTCTAAC	25
<u1* rbx*>-TAMRA	u1* rbx*	CTGAAG GTTTGACTCTAATGGTTTGGCAATG /36-TAMTSp/	31
RQ-<rbx>	rbx	/5IAbRQ/ CATTGCCAAACCATTAGAGTCAAAC	25
<u1* rby*>-TAMRA	u1* rby*	CTGAAG GTGAGACTCGTATTAGGGTGCAATG /36-TAMTSp/	31

RQ- <rby>	rby	/5IAbRQ/ CATTGCACCCTAATACGAGTCTCAC	25

Table 2.8: Strands occur in ndsDNA gates for the consensus network. Strand sequences see Table. 2.7.

Gate	**Strand**	**Bottom strand length (nt)**
Join_{XY}	Join_{XY}-Bottom, <x t>, <y t>, <rxy u1>	155
Join_{BX}	Join_{BX}-Bottom, <b t>, <x t>, <rbx u1>	155
Join_{BY}	Join_{BY}-Bottom, <b t>, <y t>, <rby u1>	155
Fork_{BB}	Fork_{BB}-Bottom, <ig>, <t b>, <u3 pb>, <t b>, <u2 rxy>	155
Fork_{XX}	Fork_{XX}-Bottom, <ig>, <t x>, <u3 px>, <t x>, <u2 rbx>	155
Fork_{YY}	Fork_{YY}-Bottom, <ig>, <t y>, <u3 py>, <t y>, <u2 rby>	155
Reporter-<u3 px>	FAM-<px* u3*>, <px>-FQ	31
Reporter-<u3 py>	TYE665-<py* u3*>, <py>-RQ	31
Reporter-<u3 pb>	ROX-<pb* u3*>, <pb>-RQ	31
Reporter-<t x>	ROX-<x* t*>, <x>-RQ	31
Reporter-<t y>	ALEX647-<y* t*>, <y>-RQ	31
Reporter-<t b>	ALEX488-<b* t*>, <b>-FQ	31
Reporter-<u1 rxy>	<u1* rxy*>-TAMRA, RQ-<rxy>	31
Reporter-<u1 rbx>	<u1* rbx*>-TAMRA, RQ-<rbx>	31
Reporter-<u1 rby>	<u1* rby*>-TAMRA, RQ-<rby>	31

Chapter 3

COMPUTING IN MAMMALIAN CELLS WITH NUCLEIC ACID STRAND EXCHANGE

Given their biocompatibility and small size, the most intriguing application of DNA nanorobots lie at the interface with biology. Applying DNA circuitry *in vivo*, however, requires special considerations to minimize unwanted interference from host cellular activity. In the work below, we demonstrated progress towards identifying design principles for adapting DNA nanotechnology to the cell.

This work's contribution is summarized as follows. Benjamin Groves, Sergii Pochekailov, Georg Seelig, and I studied the interplay between molecular structure, chemistry, and delivery method. We also developed the logic OR and AND gates. Benjamin developed the strand-exchange based siRNA. Chiara Zurla, Jonathan Kirschman, and Philip J. Santangelo developed a method to concentrate strand displacement gates by interactions with endogenous mRNA, which enables imaging of strand exchange reactions in mammalian cells. Sergii and Chiara took all the microscopy images. This work was submitted in full as:

Benjamin Groves*, Yuan-Jyue Chen*, Chiara Zurla*, Sergii Pochekailov, Jonathan Kirschman, Philip J. Santangelo, and Georg Seelig. "Computing in mammalian cells with nucleic acid strand exchange." * indicates that those authors contributed equally to this work.

3.1 Abstract

DNA strand displacement has been widely used for the *de novo* design of molecular circuits, motors, and sensors in cell-free settings. Here, we report the first steps towards adapting this technology to mammalian cells. We began by optimizing the cellular performance of fluorescent reporters based on 4-way strand exchange reactions and identified robust design

principles by systematically varying the molecular structure, chemistry and delivery method. Next, we developed and tested AND and OR logic gates based on 4-way strand exchange, demonstrating the feasibility of multi-input logic. We further established that functional siRNA could be activated through strand exchange, and we used native mRNA as programmable scaffolds for co-localizing gates and visualizing their operation with subcellular resolution. These experiments provide a first example of strand exchange-based circuitry directly interacting with a native mRNA. Together, our results suggest a practical way forward for porting the rich toolbox of DNA nanotechnology to cells.

3.2 Introduction

Biocomputers able to sense, analyze and modulate molecular information in the cellular milieu would make a valuable contribution to medicine and biological research. Dynamic DNA nanotechnology has made important progress towards the goal of building such embedded cellular controllers by first establishing systematic methods for the design of complex molecular circuits that work reliably in cell-free settings[13]. DNA realizations of Boolean logic circuits[42, 20, 6, 107], finite state machines[31], analog chemical reaction networks[35, 34], linear control systems[36], or neural networks[87] have all been demonstrated. DNA nanotechnology has also resulted in molecular sensors and amplifiers[21, 25, 41, 40, 37] that could provide inputs to such circuits, as well as molecular motors[108, 26, 109, 44, 15, 39] and switchable nanostructures[44, 110] that could be controlled by them.

Recent work has begun to demonstrate that DNA nano-devices can be compatible with complex biological environments. For example, molecular probes based on the hybridization chain reaction have enabled RNA imaging in fixed cells and tissues[45, 49]. A DNA nano-robot recognized cell surface markers on live cells and directed the delivery of a molecular payload to a subpopulation of cells[53]. A similar nano-robot was also shown to be active in the bloodstream of live cockroaches[56]. Antibody-guided DNA circuits were similarly applied for the analysis of plasma membrane targets[59]. Delivery of complex DNA nanostructures to the interior of mammalian cells has been demonstrated[111, 112]28,29, and intracellular DNAzyme-based logic gates and DNA pH sensors were shown to work reliably[62, 78]. Moreover, RNA-based structures reminiscent of DNA tiles have been expressed and assembled inside of bacteria[113]. Recent work even suggested that DNA logic gates can detect microRNA in living cells[79]. However, a systematic understanding for how to adapt DNA nanodevices to the cell is still missing.

To recreate in cells the full diversity of cell-free dynamic DNA devices there is a need to establish the design parameters that render intracellular systems as engineerable as their *in vitro* counterparts. Here, we address this challenge for DNA circuitry that relies on

strand displacement and exchange reactions, which, owing to their simplicity, underlie the vast majority of dynamic DNA nano-devices. We focus on logic gates suitable for 4-way strand exchange which minimize crosstalk with other nucleic acids in complex environments because of the predominately double-stranded nature of components[49, 74]. Gates are chemically synthesized and, like siRNAs or antisense oligonucleotides, are transiently delivered to mammalian cells rather than genetically encoded and expressed within cells. For initial characterization experiments both the logic gate and inputs are exogenous since this approach provides a degree of quantitative control over all reactant concentrations.

We seek to understand how design and delivery parameters affect gate operation (Fig. 3.1); we begin by characterizing the effect of gate architecture (i.e the length and spatial arrangement of single and double-stranded domains that make up a functional gate) and chemical composition. Then, we compare different delivery methods that permit the observation of gate activation in cells and quantitatively characterize the relationship between reaction yield and subcellular distribution of the reactants. We either use a cytosolic delivery method, via the pore-forming toxin Streptolysin O (SLO), which enabled the interaction of gates with endogenous mRNAs, or an endocytic delivery method using Lipofectamine 2000 (L2K). Finally, we demonstrate that strand exchange-based components can interface with endogenous cellular machinery, such as RISC and native mRNA, laying the foundation for future therapeutic or diagnostic applications.

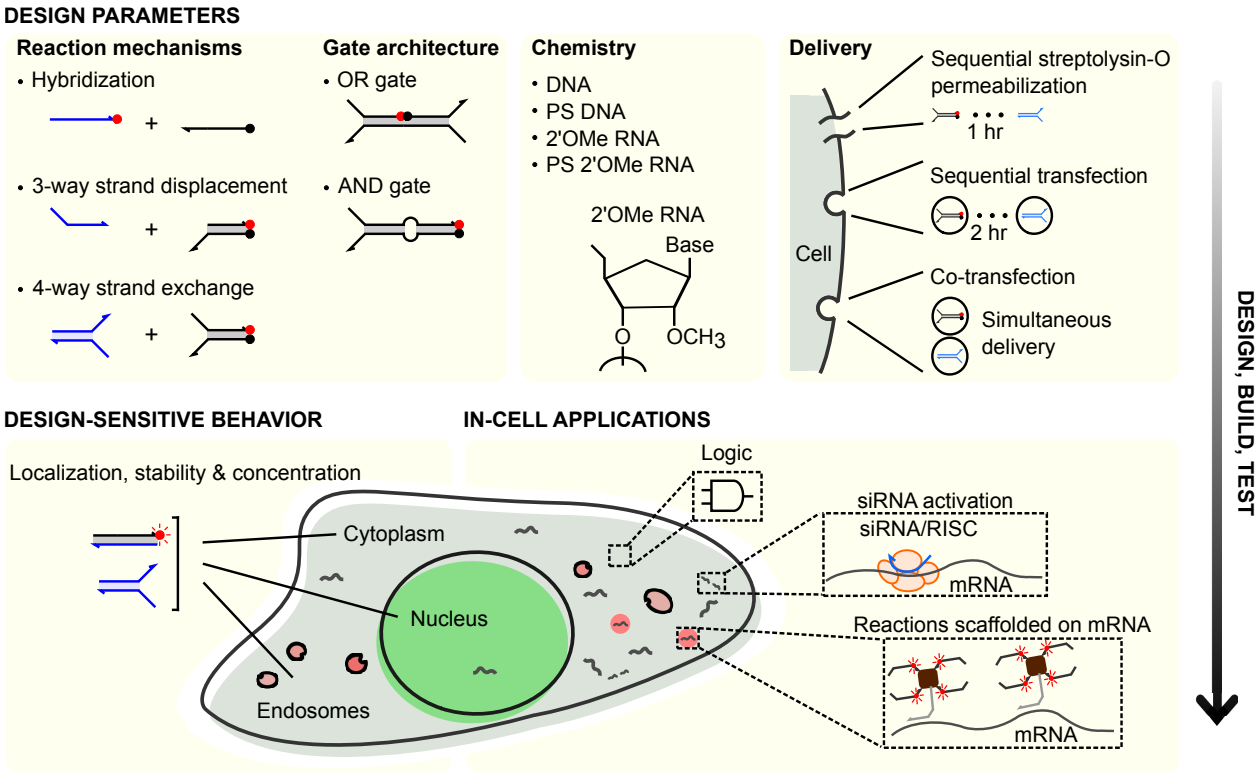


Figure 3.1: Empirical design parameters determine in-cell performance. Decisions made at the design level, such as the choice of gate architecture, nucleic acids modifications and delivery method have a strong impact on reaction kinetics, stability against nuclease degradation and subcellular localization of stand exchange-based logic gates. In turn, these properties determine how well a logic gate can fulfill its intended function including the ability to controllably interface with native mRNAs or the RNAi pathway.

3.3 4-way strand exchange mechanism and in vitro characterization.

In a 4-way strand exchange reaction (Fig. 3.4a), the two reactants (the reporter, which carried a quenched TYE665 fluorophore, and the input) are predominately double-stranded (domain a) with forked single-stranded toeholds (domains 1, 2). The single-stranded toehold domains on the input are complementary to those on the reporter (1 with 1*, 2 with 2*); hybridization of the reporter and input toeholds initiates the reaction. The reaction then proceeds via a random walk branch migration, which results in the exchange of strands between the complexes[114, 115]. Successful completion of the reaction yields two fully double-stranded complexes, one labeled with the fluorophore (the active reporter) and the other with the quencher (the waste).

We began by characterizing reaction kinetics in cell free settings. Unlike for 3-way strand displacement, we found that not only the strengths of the toehold but also the length of the double-stranded domain is a primary determinant of reaction rate (data not shown)[18]. We chose to move forward with complexes containing 6 nucleotide (nt) toeholds and 16 or 22 base pair (bp) branch migration domains, parameters that result in a good compromise between desired fast kinetics and high thermodynamic stability.

We then compared the efficiency of strand exchange with DNA probes to that of probes made from nucleic acids with phosphorothioate (PS) bonds and/or 2-O-methyl (2'OMe) ribonucleotides, whose increased stability against degradation by nucleases renders them better suited for live cell studies[116, 117, 118]. Following previous work, we limited the PS modifications to seven bases at both ends of each oligonucleotide (Table 3.2)[119]. *In vitro* measurements in high salt buffer showed that the reaction kinetics are similar with 2'OMe RNA and DNA probes, while PS bonds appeared to slow down the reaction rates (Fig. 3.2), likely because of the reduced thermodynamic stability of toehold binding interactions[120]. This difference was even more pronounced when 1x phosphate buffered saline (PBS) was used (Fig. 3.2b).

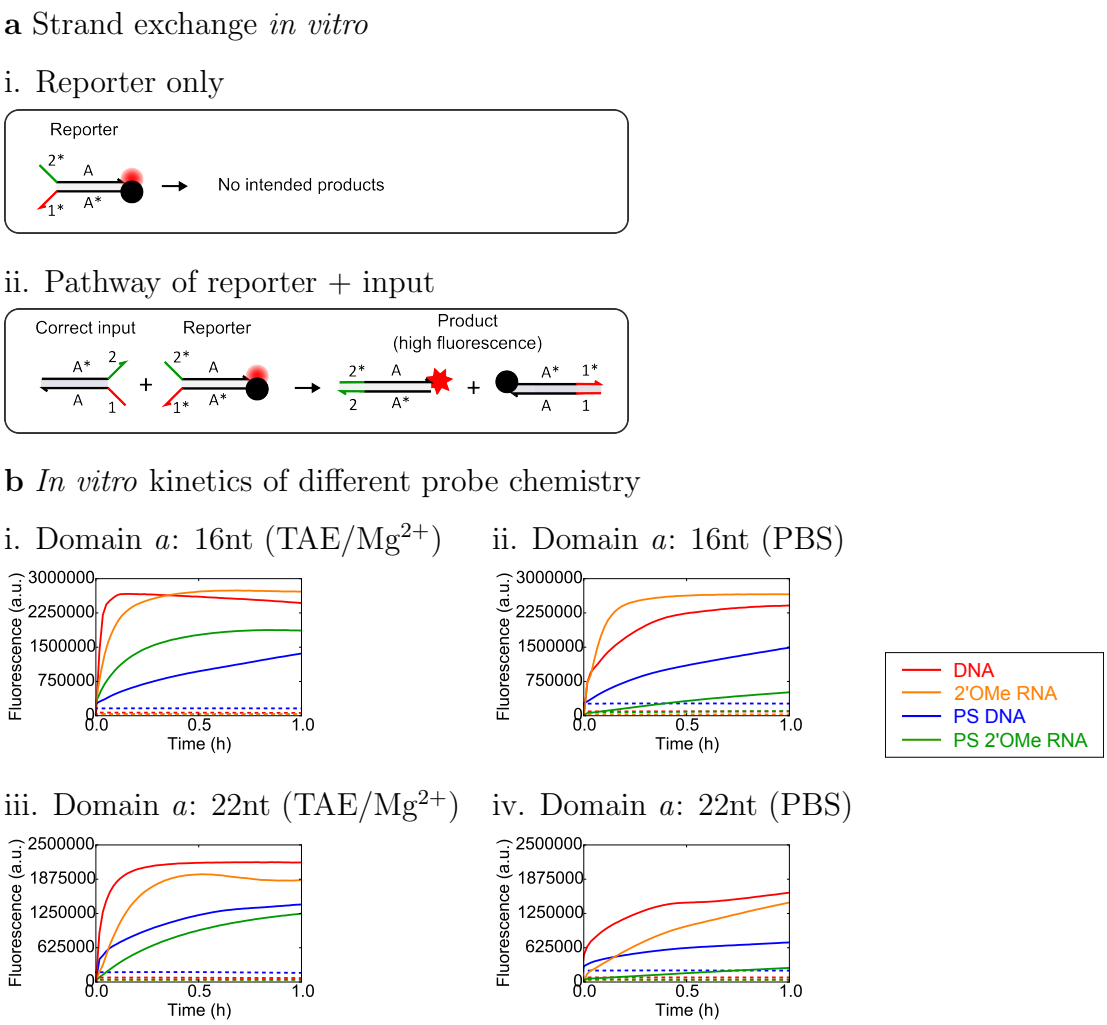


Figure 3.2: *In vitro* kinetics of 4-way strand exchange reaction. **a,** A diagram of the 4-way strand exchange reaction. Interactions between the input and the reporter result in the separation of the strand bearing the fluorophore (TYE665) from the quencher (Iowa Black RQ). **b,** *In vitro* kinetics of 4-way strand exchange reaction with different probe chemistry. Complexes with 6-nt toeholds and 16-bp double-stranded regions were mixed in 1x TAE/Mg^{2+} (i) or in 1x PBS (ii) at 37°C. Complexes with 22-bp double-stranded regions were mixed in 1x TAE/Mg^{2+} (iii) or 1x PBS (iv) at 37°C. Reporter was at 1x (50nM), and input was at 3x. Solid traces are from measurements with reporter and input; dashed traces are from measurements with reporter only.

3.4 Characterization of 4-way strand exchange in cells.

To visualize strand exchange reactions in mammalian cells, we next optimized the delivery regime. We initially focused on lipid-based transfection because this approach is experimentally straightforward and results in high cellular concentrations of nucleic acids. Moreover, to minimize the complexity of the experiment we aimed to package the reporter and input independently but deliver them to cells simultaneously. To ensure that the reporter and input did not interact before entering the cells, we tested several transfection reagents *in vitro* for their ability to stably package and insulate their nucleic acid cargo. We found that only L2K prevented the interaction between the input and reporter complex when the pre-packaged complexes were mixed in a test tube (Fig. 3.3 and Appendix 4.2)[73].

Experiments in CHO K1 cells showed that reporter activation strongly depended on the probe chemistry. A DNA-based system achieved ∼1.5 fold activation after 6 hours, and PS DNA performed only marginally better. Remarkably, replacing DNA with 2'OMe RNA made a substantial improvement, with reporter activation increasing to ∼7-fold. On the other hand, further modification of the 2'OMe RNA with phosphorothioate bonds decreased reporter triggering (Fig. 3.4b). Confocal images of CHO K1 cells fixed after the delivery of 2'OMe RNA probes revealed a similar trend (Fig. 3.4c).

Inputs with orthogonal (scrambled) toehold or double-stranded domains failed to activate the reporter (Fig. 3.5). We also modified the input with a fluorophore and removed the fluorophore from the reporter to confirm that a 4-way strand exchange reaction can be used to quench a fluorophore rather than unquench it (Fig. 3.6). Although unquenching could in principle be the result of degradation or thermodynamic instability, the success of the quenching experiments is consistent with the formation of a double stranded product. Together, these experiments suggest that the reaction indeed proceeds through the intended toehold mediated strand exchange mechanism.

As a further control, we checked that all probes were stable under endosmome-mimicking buffer conditions (data not shown). We also investigated cytotoxicity, since previous reports

have shown that some nucleic acid modifications can be toxic to cells [118, 121]. Cells remained largely viable after treatment, with the transfection reagent having the largest impact (Fig. 3.7).

Next, we tested an alternative, sequential, transfection regime. Here, the cells were first transfected with the reporter, washed, and following a two hour incubation, were then transfected with the input (Fig. 3.8a). This approach ensures that reactions can only occur within cells. In contrast, for the co-transfection approach it is likely that reactions occur at least in part during uptake. Overall, reporter activation was similar to that observed upon co-transfection, but with much reduced on/off ratios and less pronounced differences between nucleic acid chemistries (Fig. 3.8b). Below, we further explore the reasons for these differences.

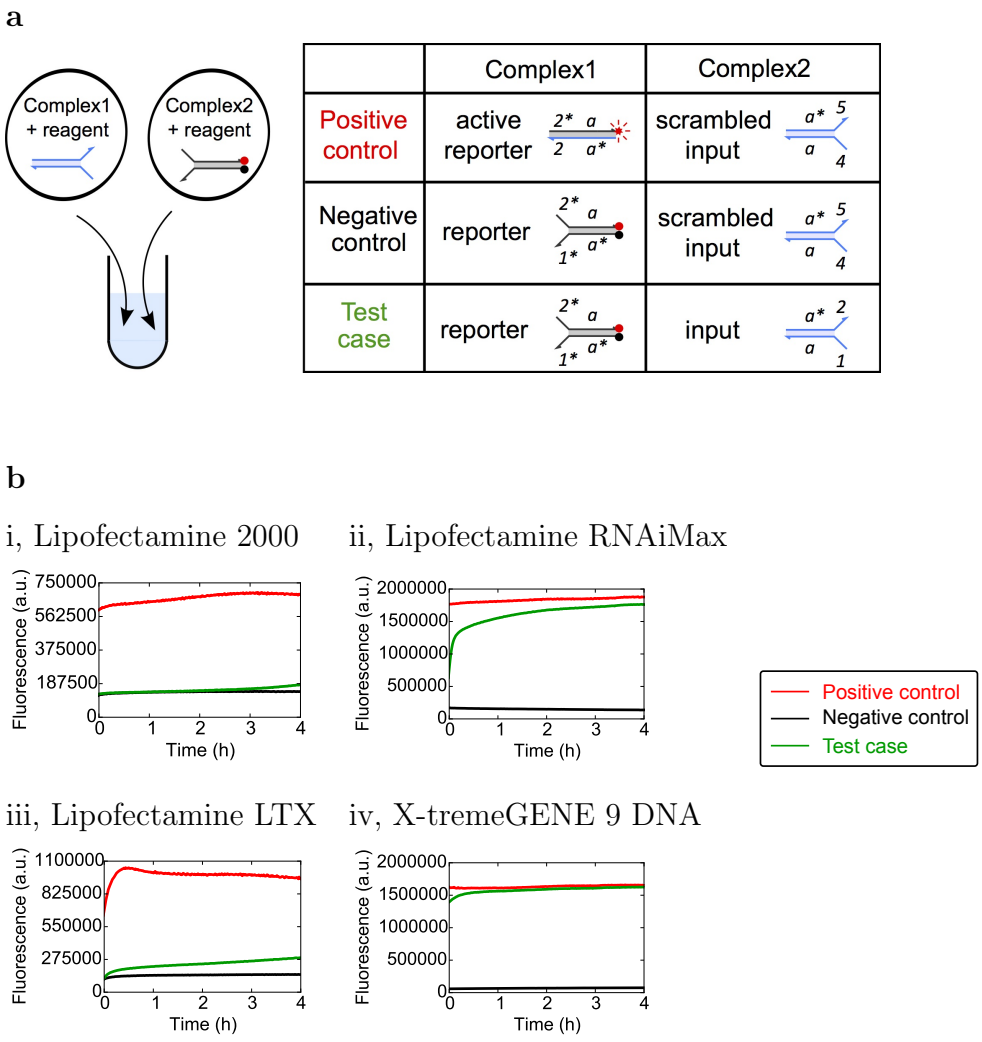


Figure 3.3: ***In vitro*** **mixing experiments. a,** Complex1 and complex2 were separately incubated with transfection reagents for 20 minutes and only subsequently mixed. The experiments were performed in Opti-MEM (a modification of Eagle's Minimum Essential Media that allows a reduction of Fetal Bovine Serum, no phenol red) at 37°C. Complex1 was at 1x (50 nM) and packaged using 2 μL transfection reagent; complex2 was at 3x (150 nM) and packaged using 3 μL transfection reagent. The complexes used in the positive control (red), the negative control (black), and the test case (green) are detailed in the table. **b,** *In vitro* kinetics of gate activation using (i) Lipofectamine 2000, (ii) RNAiMax, (iii) Lipofectamine LTX, and (iv) X-TremeGENE 9 DNA.

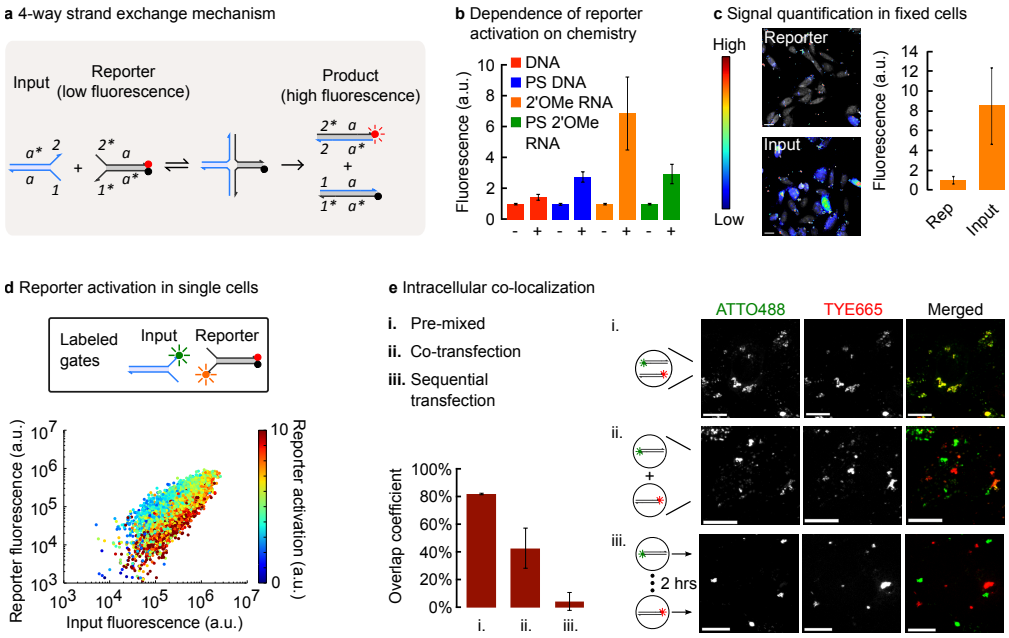


Figure 3.4: Strand exchange reactions in mammalian cells. a, 4-way strand exchange reaction mechanism. Nucleic acid strands are depicted as lines, with a half-arrow denoting the 3' end. Short, single-stranded toehold domains are numbered, while longer double-stranded domains are marked with letters. An asterisk indicates complementarity. Here the double-stranded domain is 16 bp and the toehold domains are 6 nt. 4-way strand exchange reactions are initiated when matching toeholds in different complexes hybridize. This initiation step is followed by branch migration and, finally, the formation of two new complexes. The reaction results in the unquenching of a fluorophore (TYE665). **b,** Dependence of reporter activation on chemistry. Reporter and input were co-transfected into CHO K1 cells, and fluorescence was measured using flow cytometry. The reporter was transfected at 1x (50nM) and the control RNA or input at 3x. For each chemistry, the fluorescent values of reporter & input (+) are normalized to the inactive reporter & control RNA (-). Error bars represent the standard deviation of at least two experiments. **c,** Representative confocal images of 2'OMe RNA reporter activation in fixed CHO K1 cells. Signal intensity is represented by color, with blue signifying low intensity and red high intensity. Total fluorescent signal was measured using confocal images of at least 10 cells from each condition. The signal was normalized to the reporter & control RNA data. Scale bars are 15 μm. **d,** Input and reporter concentrations were tracked using unquenched fluorophores (ATTO488 and TYE563, respectively). The signal intensity of the input and reporter gates from cells transfected with 1x of both gates is plotted along the x- and y-axes. The signal from reporter activation is normalized to the reporter signal (i.e. TYE665/TYE563) to give a measure of the efficiency of reporter activation. **e,** Two non-interacting 2'OMe RNA complexes labeled with TYE665 (red) and ATTO488 (green) were mixed before being incubated with L2K in the same tube (Pre-mixed; i.), mixed with L2K independently (Co-transfection; ii.), or mixed with RNAiMAX independently and added to cells separately with a two hour gap (Sequential transfection; iii.). Confocal images of fixed cells were used to determine the overlap coefficient of the two complexes in all conditions. Scale bars are 10 μm. Error bars are standard deviation.

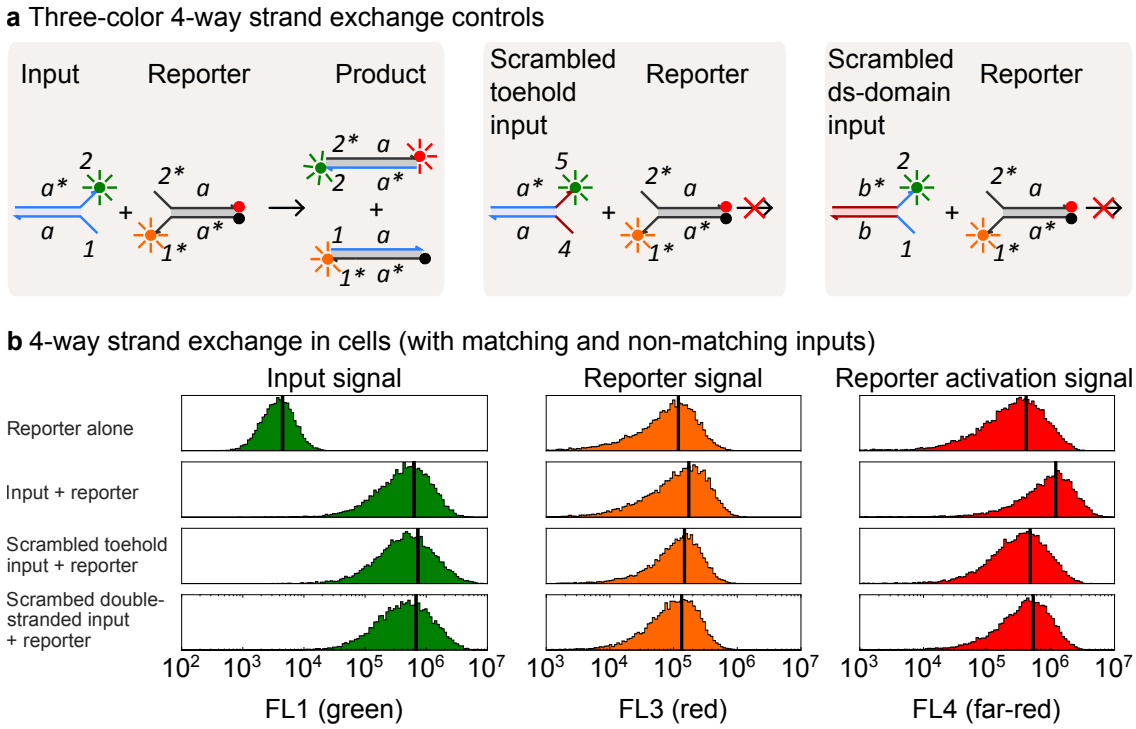


Figure 3.5: Strand exchange controls (scrambled domains). a, (from left to right) Schematic of 4-way strand exchange with fluorescently labeled input and reporter, an input with orthogonal toeholds or an input with an orthogonal double-stranded domain. Strand sequences see Table 3.3. **b,** Histograms of cell count versus fluorescent signal for the indicated channels. The data was processed as described in section 3.12.13. The black lines represent the mean value. The indicated gates (or control RNA) were co-transfected into cells via Lipofectamine 2000 transfection (as described in the materials and methods section).

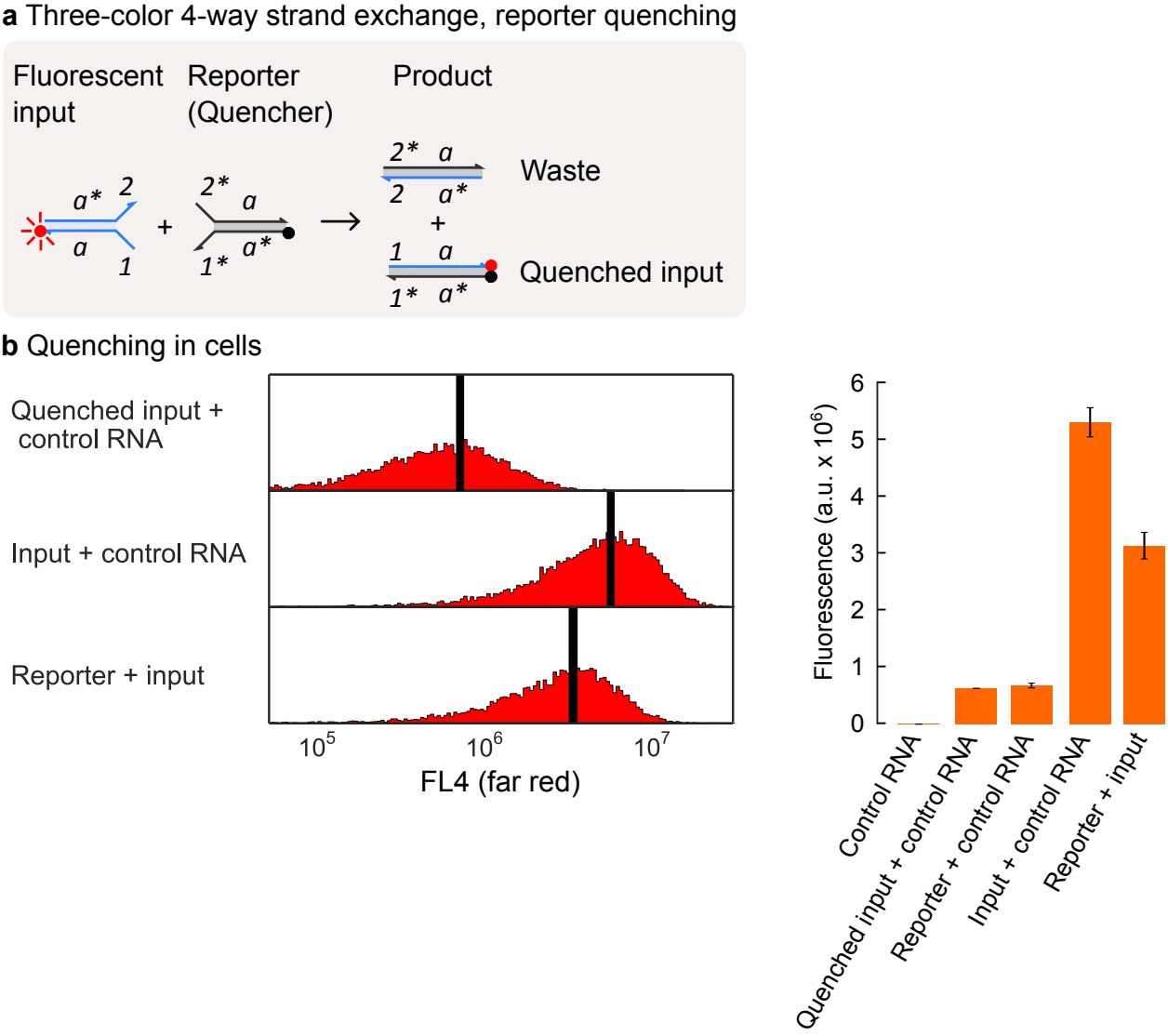


Figure 3.6: 4-way strand exchange quenching reaction. **a,** Schematic of 4-way strand exchange gates. The fluorescent signal is quenched following the interaction of the fluorescent input and the reporter labeled with a quencher moiety. **b,** Histograms of cell count versus fluorescent signal for the indicated channels. The data was processed as described in section 3.12.13. The black lines represent the mean value. The indicated gates (or control RNA) were co-transfected into cells via Lipofectamine 2000 (as described in the materials and methods section).

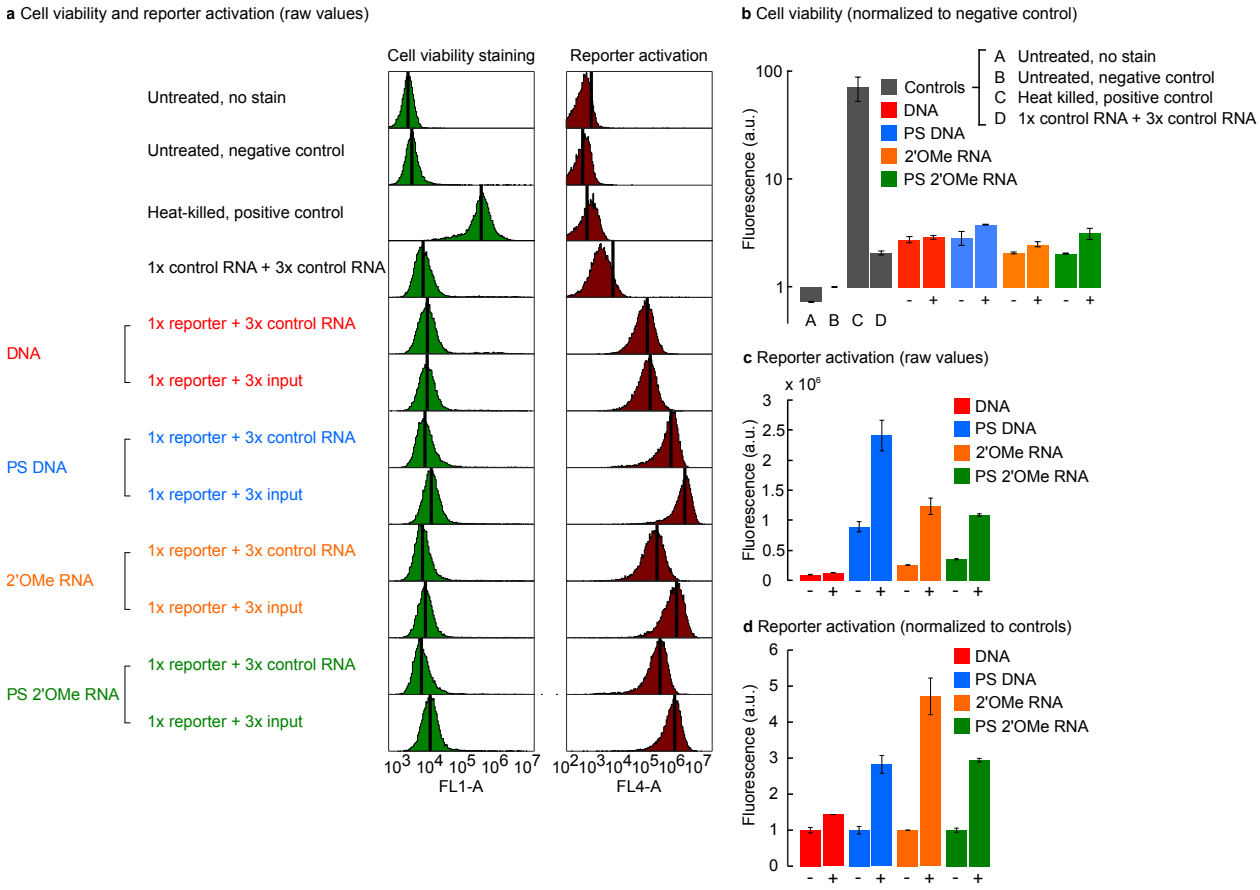


Figure 3.7: Cell viability following transfection of gates. a, Cells were co-transfected with gates composed of the indicated nucleic acid chemistry for four hours and then incubated in regular growth media for an additional two hours. Positive control: cells that were heat-killed were treated with trypsin and then incubated in 1x PBS + 2% FBS at 65 °C. We measured both the cell viability (FL1 channel-green), and the reporter activation (FL4-far-red). The black lines indicate the median and mean for the green and far red signals, respectively. **b,** Cell viability: median intensity of fluorescence in each condition. All samples are normalized to the untreated, stained control. **c and d,** Average reporter activation from the same set of experiments. **c,** Raw fluorescence value. **d,** Normalized fluorescence value (data is normalized to the inactive reporter for each nucleic acid chemistry). Error bars represent the standard deviation of at least two experiments.

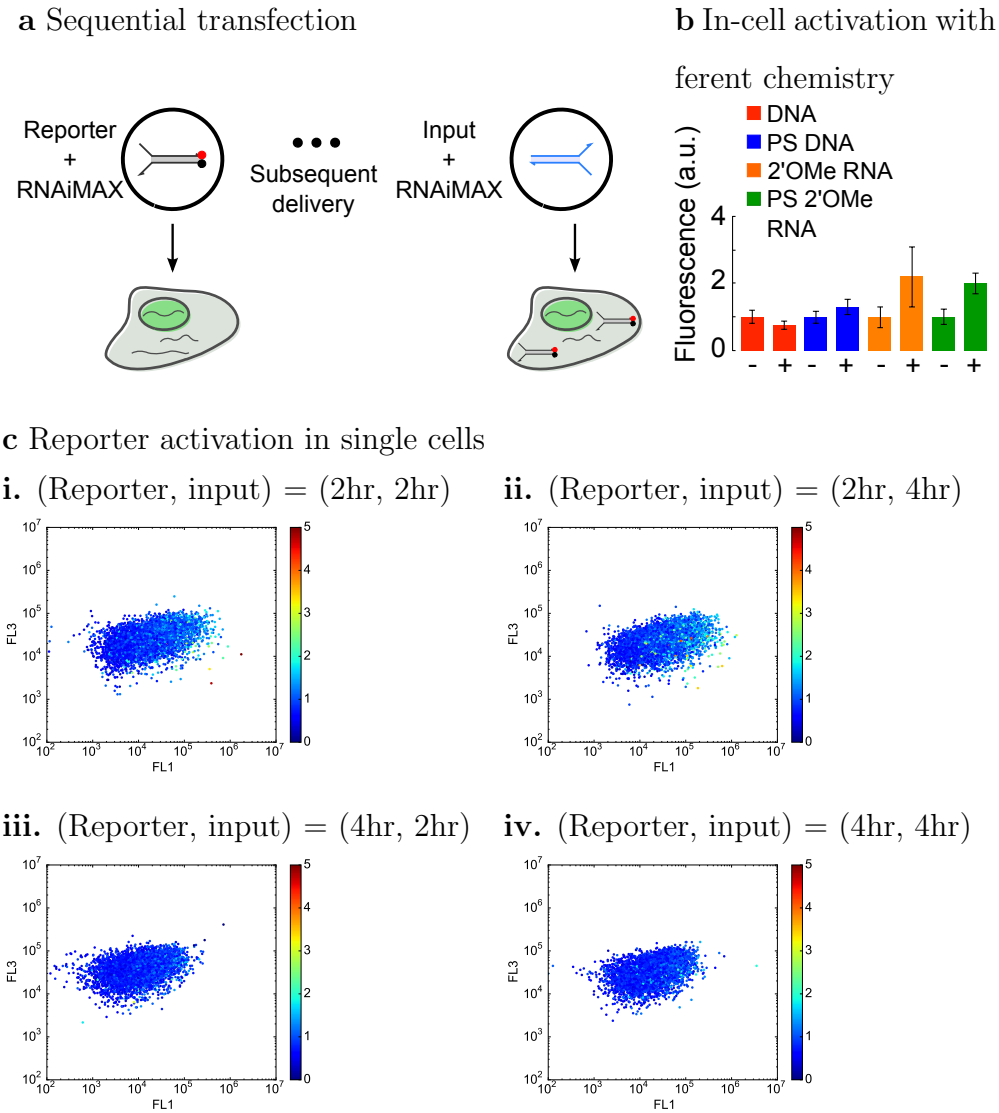


Figure 3.8: Sequential transfection. **a,** Sequential transfection method. **b,** Reporter activation in CHO K1 cells. Reporter was at 1x (50nM), and input was at 3x. Fluorescence in cells was measured using flow cytometry. For each chemistry, fluorescence in cells was measured using flow cytometry and normalized to the signal of the inactive reporter (-). Error bars represent the standard deviation of at least two experiments. **c,** Overlap of fluorescently labeled reporter and input following sequential transfection. The reporter and input were labeled with ATTO488 and TYE563, respectively, and transfected into CHO K1 cells at 50 nM using Lipofectamine RNAiMAX (Life Technologies). The reporter was first delivered to the cells, and allowed to incubate for 2 or 4 hours. The cells were washed twice with 1x PBS. Then the input was delivered to cells, allowing incubation for another 2 or 4 hours as indicated on top of each panel. Cells were washed twice with 1x PBS, and MEM/alpha media + 10% FBS was added for a final two hour incubation. The signal from reporter activation is normalized to the reporter signal (i.e. TYE665/TYE563) to give a measure of the efficiency of reporter activation.

3.5 Impact of the inter- and intra-cellular distributions of the reporter and input.

In both CHO K1 and HeLa cells, adding an equal molar ratio or even an excess of input never fully triggered the reporter (data not shown). We hypothesized that this apparently incomplete activation is due to an uneven distribution of input and reporter among cells, such that in many cells there is insufficient input to activate all of the reporter gates. To track the intracellular concentrations of reporter and input we added a constitutively active fluorophore to both complexes (TYE563 and ATTO488, respectively) and co-transfected cells as described. We found that cells with a higher input:reporter ratio exhibited an increased efficiency of reporter activation, calculated by dividing reporter activation by reporter concentration (Fig. 3.4d). The same trend was seen when the reporter and input were transfected sequentially (Fig. 3.8c), again suggesting that the degree of input and reporter co-localization is a major determinant of reporter activation.

The efficiency of reporter activation may be further limited by heterogeneities in the subcellular distributions of the gates. When analyzing the reporter signal in confocal images, signal was observed in the cell nucleus, cytoplasm and endosomal granules for both 2'OMe RNA or PS 2'OMe RNA reporters, but we noted that the distribution was not uniform (data not shown). Thus, even if the reactants are present in equal amounts in a given cell, they can interact only to the extent that their distributions overlap.

To further investigate the degree to which the distributions of two independent complexes overlapped in the same cell, we measured the co-localization (Manders overlap coefficient) of two orthogonal fluorescent duplexes in CHO K1 cells. Each duplex carried a constitutively active fluorophore (ATTO 488 or TYE 665) and was delivered by co- or sequential transfections. As a control, we transfected duplexes that had been incubated together before being mixed with L2K. Co-localization of the complexes was strongly influenced by the transfection method, decreasing from 40% in co-transfections, to 5% in the sequential transfections (Fig. 3.4e). These results give a physical explanation for the differences we observed between

the two delivery methods and also highlight an important consideration specific to cell-based experiments.

3.6 OR and AND logic in cells.

To demonstrate that more complex strand exchange systems also function in cells we designed logic OR and AND gates based on 4-way strand exchange (Fig. 3.11a). The OR gate consists of three strands, a long bottom strand and two shorter top strands. Dye and quencher labels are attached to the short top strands such that the labels are co-localized at the center of the structure when the gate is intact. The forked toeholds that appear on either side of the gate engage with either input A or B in a 4-way strand exchange reaction; input A displaces the fluorophore strand from the OR gate, while input B displaces the quencher strand. Both reactions thus can result in high fluorescence. In a cell-free setting, the reaction mechanism worked as designed for gates made of either 2'OMe RNA or PS 2'OMe (Fig. 3.9).

Upon testing in cells, we found that an OR gate made entirely of 2'OMe RNA did not appear to exhibit the expected behavior. Specifically, addition of input B did not result in high final fluorescence. We hypothesized that the fluorescent end product of that reaction, a nicked 38 base pair duplex with forked overhangs on one end only, was susceptible to degradation by cellular nucleases that bypass shorter duplex domains. To solve this problem, we experimented with OR gates and inputs made up of a combination of 2'OMe RNA strands with and without PS bonds. Although all of the combinations behaved as expected *in vitro*, in cells the complexes that predominantly consisted of PS 2'OMe RNA strands were less efficiently activated, but appeared to have more uniform stability (Fig. 3.10). We decided to change one of the strands of the input B to PS 2'OMe RNA, increasing the stability of the longer double-stranded product while also allowing the OR gate to be activated 2.5-3 fold by either input alone or both together (Fig. 3.11a).

The AND gate is made of two strands and has a set of forked toeholds at the end opposite of the dye and quencher label (Fig.3.11b). Integral to the gate design is the 'hiding' of the second set of toeholds (1* and 2*) in looped DNA. Two inputs must interact with the AND gate consecutively to trigger fluorescence. Only following the interaction of the first signal (input B) with the gate are these made available for binding to the second signal (input A).

We measured AND gate (implemented using 2'OMe RNA) activation in cells using flow cytometry (Fig. 3.11b). Addition of both inputs resulted in ∼3.5-fold activation, while only negligible activation occurred when only one input was used (Fig. 3.11b).

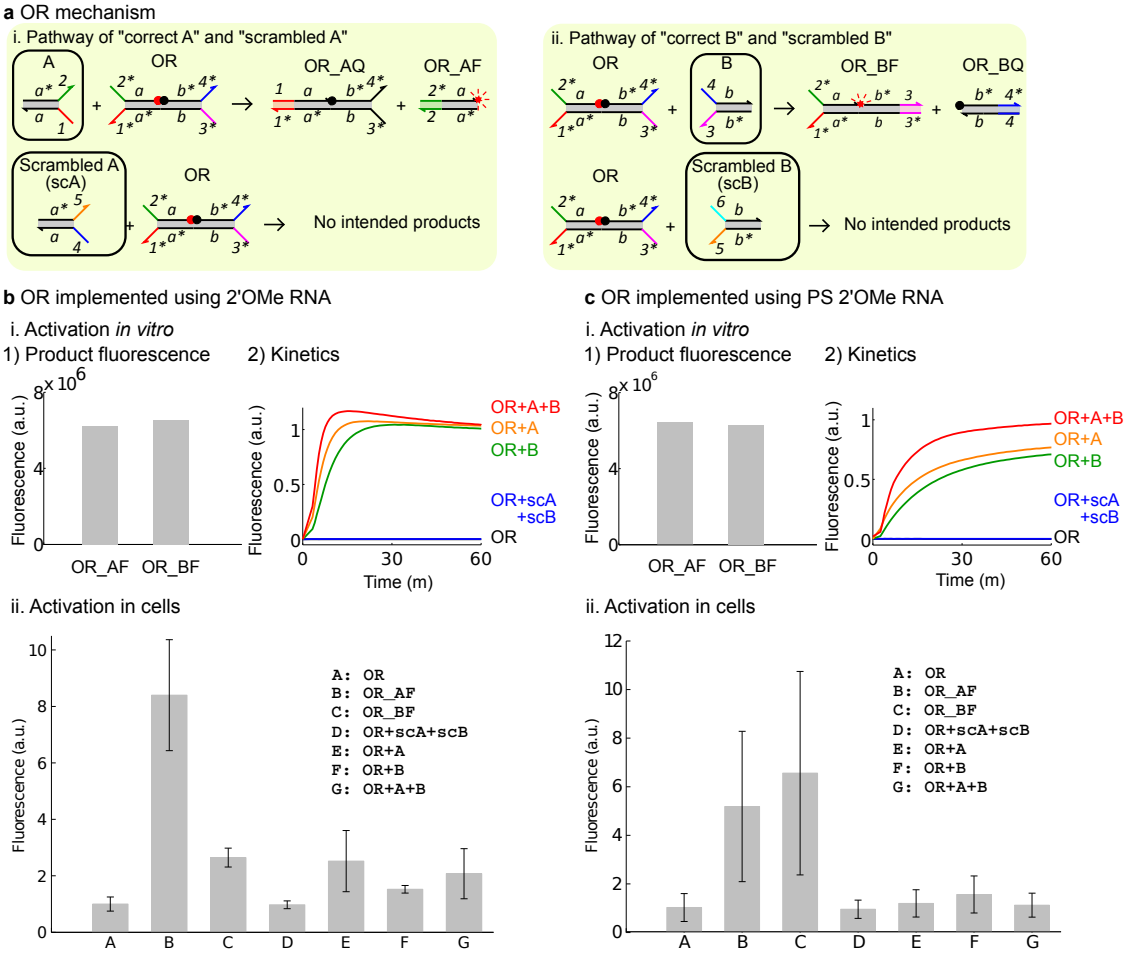


Figure 3.9: OR logic implemented using either 2'OMe RNA or PS 2'OMe RNA. Input A can interact with the OR gate to release OR_AF and OR_AQ complexes; input B can interact with the OR gate to release OR_BF and OR_BQ complexes. The OR gate does not have intended pathway for spurious inputs, scrambled A (scA) and scrambled B (scB). *In vitro* experiments of the OR gate were performed in 1x TAE/Mg^{2+} at 37°C. The OR gate was at 1x (50nM); all inputs were at 3x. Panel (i) shows the activation *in vitro*: 1) the fluorescence of two pre-annealed products, OR_AF and OR_BF. 2) the time-course data of the OR gate using different inputs. Panel (ii) shows the activation in CHO-K1 cells. Fluorescence was normalized to the signal of inactive reporter. Error bars represent the standard deviation of at least two separate experiments. **b**, OR logic (the OR logic gate and all inputs) implemented using 2'OMe RNA alone. **c**, OR logic (the OR logic gate and all inputs) implemented using PS 2'OMe RNA alone.

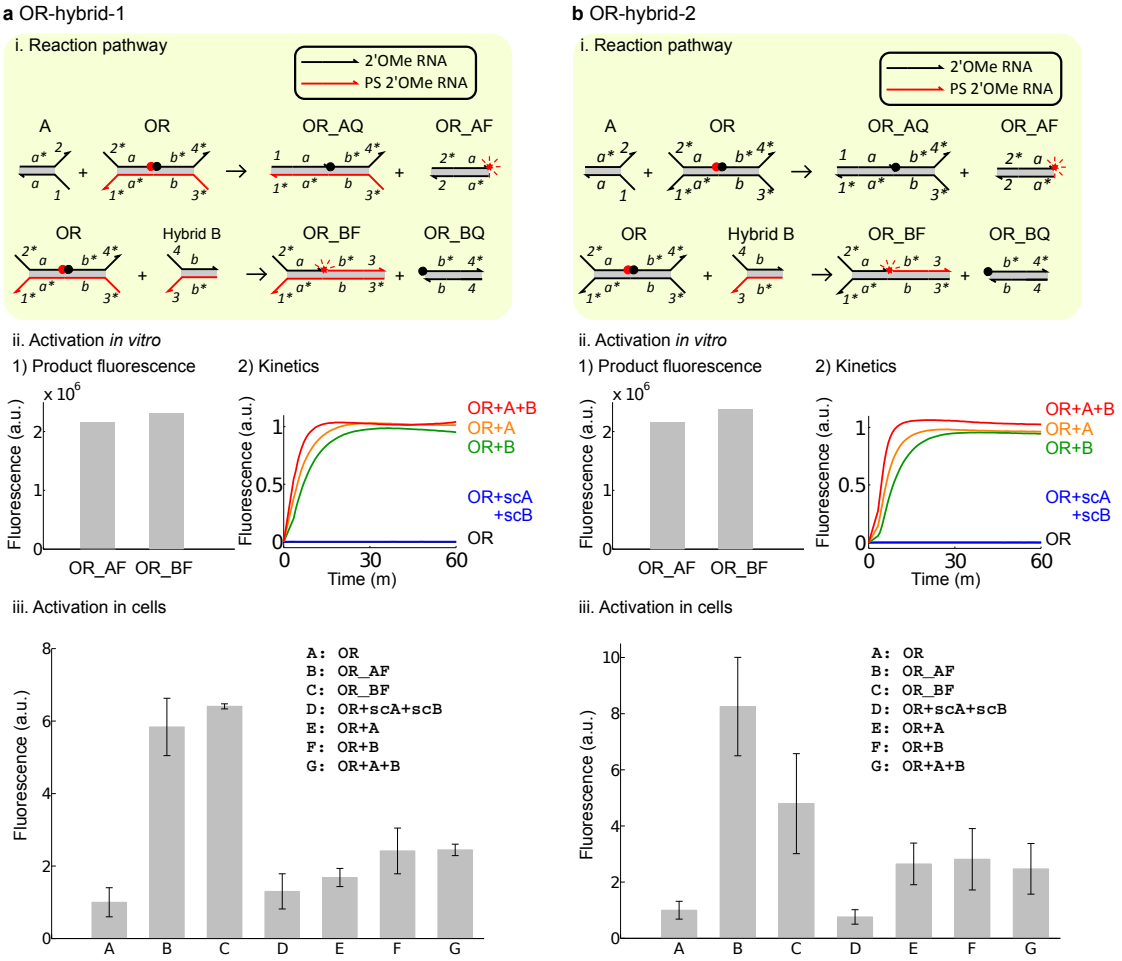


Figure 3.10: Hybrid implementation of OR logic. Panel (i) shows the reaction pathway and components of two different OR hybrid systems **a,** OR-hybrid-1: OR, hybrid-B, and OR_BF were hybrids of PS 2'OMe RNA and 2'OMe RNA strands; other complexes were made of 2'OMe RNA strands alone. **b,** OR-hybrid-2: hybrid-B and OR_BF were hybrids of PS 2'OMe RNA and 2'OMe RNA strands; other complexes were made of 2'OMe RNA alone. *In vitro* experiments of the OR gate were performed in 1x TAE/Mg^{2+} at 37°C. The OR gate was at 1x (50nM); all inputs were at 3x. Panel (ii) shows the activation *in vitro*: 1) the fluorescence of two pre-annealed products OR_AF and OR_BF. 2) the time-course data of the OR gate using different inputs. Panel (iii) shows the activation in CHO-K1 cells. Fluorescence was normalized to the signal of inactive reporter. Error bars represent the standard deviation of at least two separate experiments.

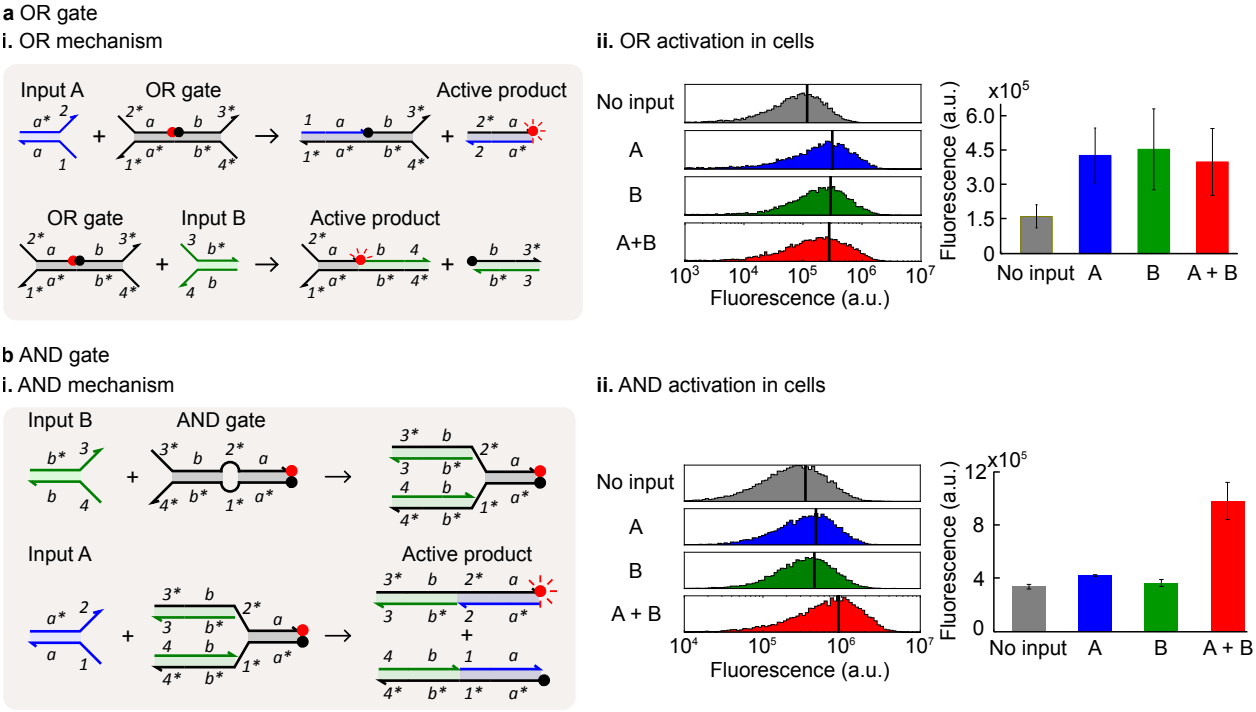


Figure 3.11: Strand exchange-based OR and AND logic gates work in mammalian cells. a, OR logic gate, **b,** AND logic gate. i. Gate design and reaction schematic. ii. Flow cytometry results. Cells were co-transfected with the indicated logic gate at 1x (50nM) and input (or control RNA) at 3x. The fluorescence signal was measured using flow cytometry. Left: Each histogram displays the cell counts versus fluorescence due to fluorophore labeling on the logic gate (the black line indicates the mean value). The inputs are indicated on the left, where less than two inputs are indicated a control RNA complex was added as well. Right: bar graphs of average fluorescence. Inputs are indicated on the x-axis. Error bars represent the standard deviation of at least two separate experiments.

3.7 Activation of an siRNA through strand exchange.

Next, we demonstrated that strand exchange reactions could generate RNA substrates recognized by the endogenous RNAi machinery. In cells, such substrates originate as double-stranded structures, and siRNA activity is sensitive to both the size of the duplex and its chemical composition[118, 122, 123, 124]. We designed a 4-way strand exchange reaction that yields a double-stranded active siRNA from two inactive complexes. To optimize the design of the initial (inactive) and final (active) complexes, we built a GFP-based sensor for siRNA activity and tested RNA-RNA, RNA-DNA, RNA-PS DNA, RNA-2'OMe RNA and RNA-PS 2'OMe RNA duplexes for their ability to knock down GFP (Fig. 3.12a). Other than RNA, none of the other nucleic acids were able to serve as fully functional guide strands (Fig. 3.12b); importantly, only when the passenger strand was PS 2'OMe RNA was siRNA activity completely lost (Fig. 3.12c).

Based on these results, the inactive starting complexes were composed of an RNA-PS 2'OMe RNA hybrid. Following a 4-way strand exchange reaction was a double-stranded RNA product formed that could enter the RNAi pathway (Fig. 3.12d). CHO K1 cells were co-transfected with the inactive starting complexes and 24 hours later, they were transfected with the reporter plasmid. Cells were analyzed 48 hours post-transfection via flow cytometry (for analysis details see section 3.12.15). The siRNA activated through strand exchange was approximately as efficient in knocking down GFP as a pre-assembled siRNA of the same sequence, reducing the fluorescent reporter signal by more than 60% (Fig. 3.12e). Activation of an siRNA can thus form a useful readout of strand exchange reactions in cells.

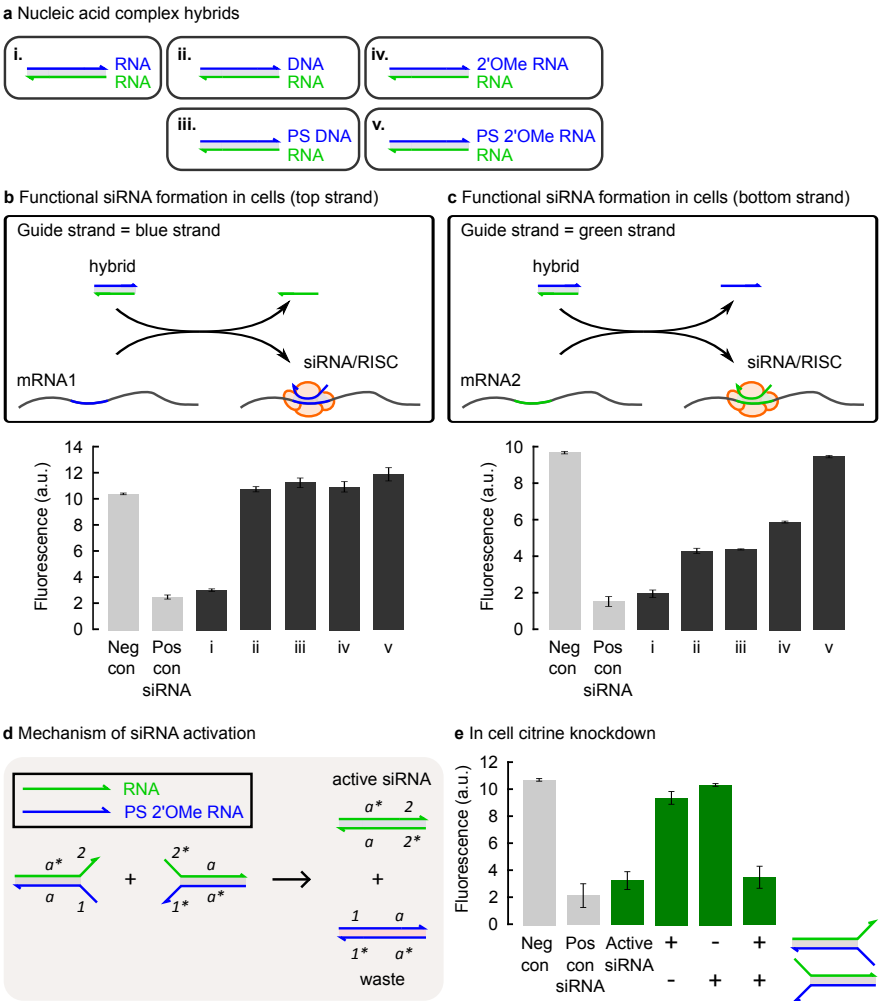


Figure 3.12: A functional siRNA can be activated through 4-way strand exchange. a, Nucleic acid hybrids were investigated for their ability to enter the RNAi pathway. Putative "siRNAs" were tested against two sensor plasmids, one sensitive to the top blue strand (**b**) and the other to the bottom green strand (**c**). The columns in both graphs correspond to cells treated with a negative control siRNA, a positive control anti-GFP siRNA, the RNA duplex (i), a DNA/RNA duplex (ii), a PS DNA/RNA duplex (iii), a 2'OMe RNA/RNA duplex (iv), or a PS 2'OMe RNA/RNA duplex (v). **d,** Each initial complex is a hybrid of an RNA strand (green line) and a PS 2'OMe RNA strand (blue line). Following strand exchange, a functional siRNA is formed. **e,** CHO K1 cells were transfected with a negative control siRNA, a positive control anti-GFP siRNA, a pre-assembled active siRNA, either one of the two hybrid PS 2'OMe RNA/RNA complexes (along with a control siRNA) or both hybrid complexes. Twenty-four hours later, the cells were transfected with a plasmid encoding mCitrine bearing a target sequence for the siRNA in the 3'UTR. mCitrine fluorescence was assayed using flow cytometry after ∼48 hours. Error bars represent the standard deviation of at least two experiments.

3.8 Use of endogenous mRNA as a scaffold for strand exchange reactions.

Up to this point, we have described methods that measure the interactions between large numbers of gate complexes. To observe reactions that occur predominately in the cytosol between smaller numbers of complexes, we next adapted our fluorescent strand exchange system to a method first developed by Santangelo *et al.*, which relies on interactions with endogenous mRNAs to concentrate weak fluorescent signals[68]. The method is based on the use of multiply-labeled, tetravalent imaging probes (MTRIPs), which have been previously demonstrated to target endogenous mRNAs in living cells[68, 125]. Native mRNA can be long lived (up to 40 hours), allowing probes time to react, and they are often aggregated into large mRNP complexes, on the order of 100 nm, concentrating the reactants within the cytosol, making reaction products easier to discern from background. Thus, endogenous mRNAs can be used as scaffolding for nucleic acid-based circuits to facilitate and visualize localized hybridization and strand exchange experiments within the cytosol of cells.

We modified the original probe design to create monovalent MTRIPs (mMTRIPs), consisting of a neutravidin core combined with two types of 2'OMe RNA oligos, one that targets endogenous mRNAs, and the other modeled on the above strand exchange systems. The "targeting" oligo complementary to an endogenous mRNA sequence is covalently bound to the neutravidin protein core; the four reporter gates are attached via the biotin-neutravidin linkage (Fig. 3.13a). mMTRIPs were delivered using reversible permeabilization mediated by the pore forming toxin streptolysin-O (SLO) and activated with a separately delivered input, also via SLO. An advantage of this delivery method is the low level of cytotoxicity, making it amenable to sequential rounds of delivery without compromising cell viability, altering cell morphology or causing observable toxic effects[68, 126].

We first compared the ability of mMTRIPs to target endogenous β-actin (ACTB) mRNA, c-myc mRNA or poly-adenylated (polyA+) mRNAs in A549 cells with that of the classical MTRIPs, and chose ACTB mRNA as a scaffold to observe reporter activation (data not shown). Using SLO, we delivered mMTRIPs targeting five separate sequences of ACTB

mRNA to A549 cells followed by either 1x or 3x of the input. We initially used this approach to observe the effect of sequential deliveries in a simple capture assay, where fluorescently labeled single strands used in place of the reporters were quenched approximately 7-fold by the addition of a complementary oligonucleotide carrying the quencher (Fig. 3.14). Then we measured reporter activation in a 4-way strand exchange system, where addition of inputs increased reporter signal by 2-fold (Fig. 3.13b). Inputs with scrambled toehold domains failed to elicit a response (Fig. 3.14). Importantly, when neutravidin-reporter complexes lacking the mRNA-targeting oligos were delivered to cells, no statistically significant activation was detected (Fig. 3.13c).

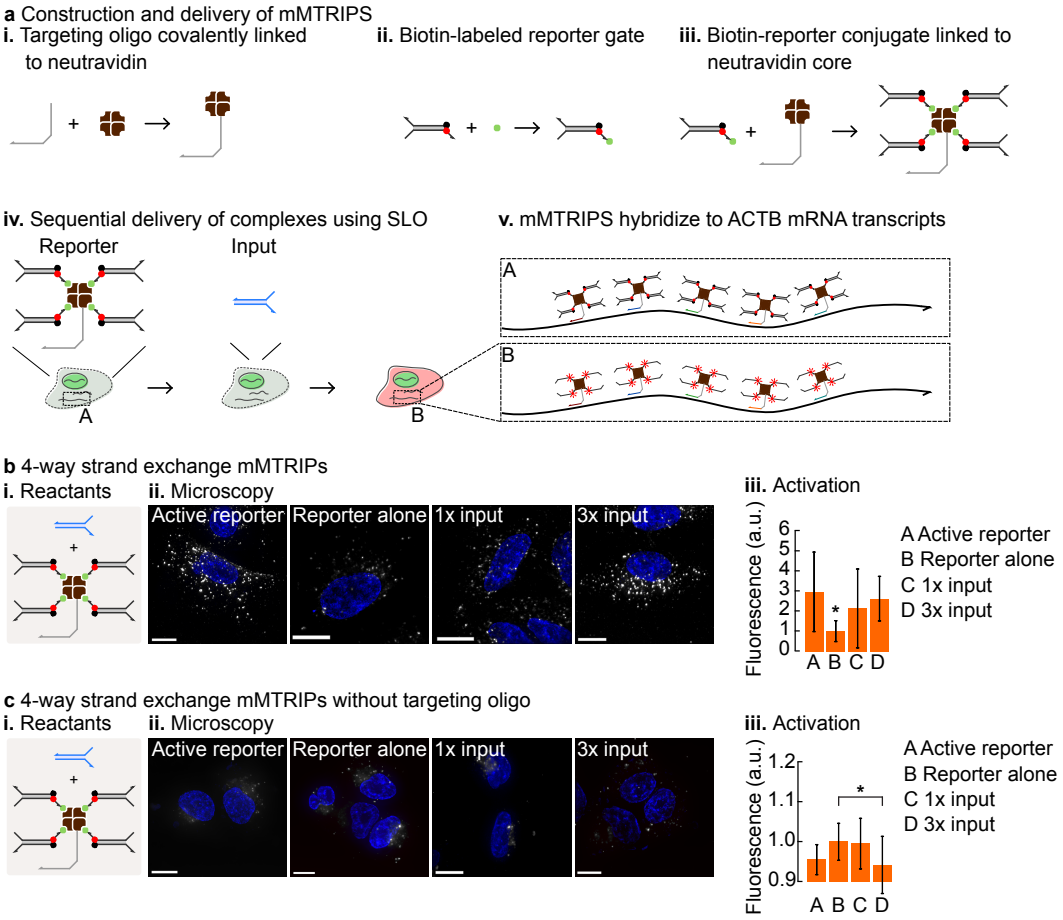


Figure 3.13: Using endogenous mRNA and multiply-labeled, tetravalent imaging probes (mMTRIPS) as scaffolds for strand exchange reactions. a, i. An mRNA targeting oligo is covalently linked to NeutrAvidin (Sigma) via an aromatic hydrazine and aldehyde linkage (Hynic-4FB, Solulink), with neutravidin naturally forming tetramers. ii. One oligo making up the 2'OMe RNA reporter has an additional poly-T linker and biotin moiety at the 5' end. Note that to increase brightness and photostability for the microscopy measurements, Cy3b/BHQ2 fluorophore/quencher pair was utilized instead of the TYE665/IAB. iii. The biotin-labeled gate associates with the NeutrAvidin tetramer to form a multiply-labeled, tetravalent imaging probe (mMTRIP). iv-v. mMTRIPS targeting five separate sites on the β-actin mRNA (ACTB) were introduced into A549 cells using SLO delivery (15 nM each). Following an hour-long recovery period, the indicated input complex was introduced at a 1x or 3x molar excess (75 or 225 nM, respectively) using SLO delivery. 4-way strand exchange reactions were visualized using mMTRIPs scaffolded on endogenous ACTB mRNAs (**b**) and un-targeted mMTRIPs (**c**). i. Details of the reactants. ii. Representative epifluorescence microscopy images. Images were deconvolved and analyzed to quantify the fluorescence signal of individual puncta and the mean fluorescence intensity was determined. iii. Mean fluorescence intensity, normalized to the quenched reporter. * indicates statistically different data according to Kruskal-Wallis One Way Analysis of Variance on Ranks. Scale bars are 10 μm. Error bars indicate one standard deviation.

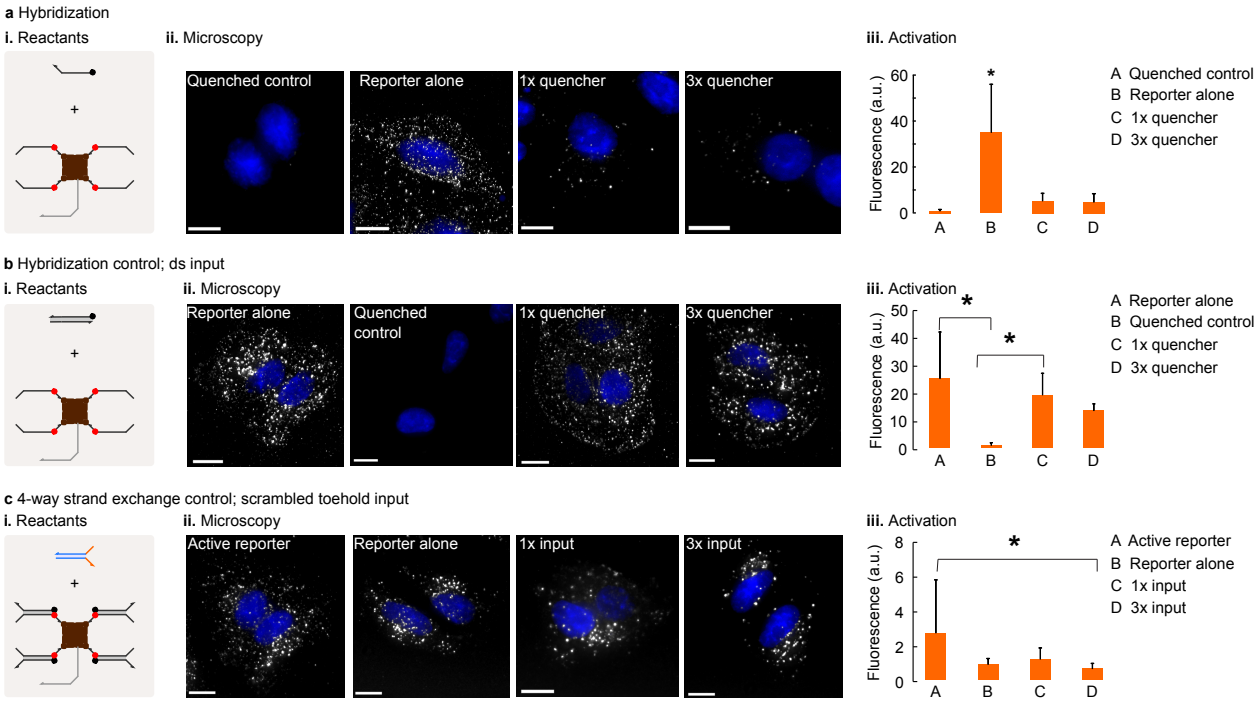


Figure 3.14: Hybridization and strand exchange reaction negative controls. The indicated experiments appear in a-c. i. Depiction of the control gates. ii. Representative images of each condition. iii. Quantification of the fluorescence intensity normalized to the signal for the quenched reporter (Q). * indicates statistically different data according to Kruskal-Wallis One Way Analysis of Variance on Ranks. Nuclei stained using DAPI appear in blue.

3.9 Scaffolding AND logic on endogenous mRNA.

Since mMTRIPs proved useful for detecting and facilitating strand exchange reactions within the cytosol, we next tested whether this approach was compatible with the AND gate design described above. We used mMTRIP-AND gates targeting three sites on endogenous ACTB mRNA. Only in the presence of both inputs A and B did we observe a 2 fold activation of the AND gates, while only negligible activation occurred when a single input was used or in the presence of inputs with scrambled toeholds (Fig. 3.15). The use of 1x molar excess of inputs was not sufficient to detect reporter activation (data not shown). It should also be noted that the incubation time following the second delivery had to be reduced to 30 minutes to limit non-specific activation of the reporter.

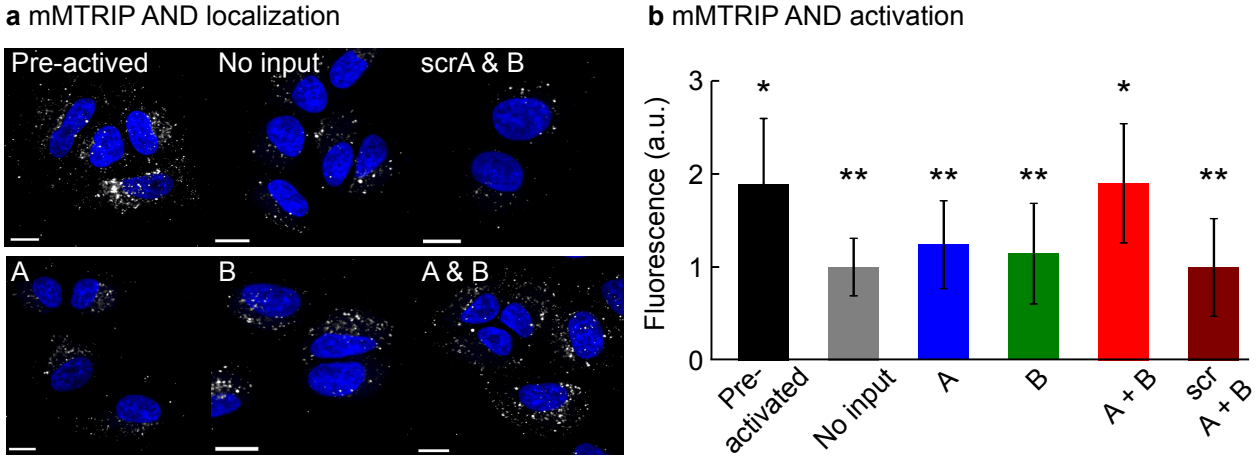


Figure 3.15: ACTB mRNA-scaffolded mMTRIP AND logic in cells. Cells were first transfected using SLO with three mMTRIP probes targeting three distinct sites on the ACTB mRNA, they were subsequently transfected with the indicated input species (scr stands for scrambled toehold). **a,** Representative confocal images of the mMTRIP AND logic gates in cells. **b,** Quantification of the signal from confocal microscopy images. The values of bars marked with a single asterisk are significantly different from those marked with two asterisks as determined by a Kruskal-Wallis One Way Analysis of Variance on Ranks. Scale bars are 10 μm. Error bars are standard deviation. Microscopy data was collected from at least 10 cells.

3.10 Discussion.

In this work, we established design principles and experimental approaches that provide a path for adapting to the cellular environment the large diversity of dynamic strand displacement devices previously tested in cell-free settings. We focused on the 4-way strand exchange mechanism and systematically explored how the interplay of molecular structure, chemistry and delivery method determines the performance of strand exchange-based logic gates and reporters in cells.

We found the chemical composition of the nucleic acid complexes to be a key determinant of system function. Among the modifications tested, 2'OMe RNA resulted in the highest degree of gate activation while also having minimal cytotoxicity. Moreover, we demonstrated that hybrid complexes consisting of RNA and PS 2'OMe RNA strands could be used to control the activation of a siRNA through strand exchange, thus connecting a bio-orthogonal system with a signal that is interpretable by the cell. However, we also observed that the interplay between gate structure and chemical modifications could affect performance in unforeseen ways. For example, we had to integrate PS modifications into one input used in the OR gate experiments in order to ensure correct operation. Thus, although the systematic use of 2'OMe RNA provides a good starting point for building strand exchange systems that function in cells, additional optimization is required on a case-by-case basis in order to properly control interactions with the cell.

The delivery method strongly influenced performance by controlling the distribution of nucleic acid devices between and within cells. Co-delivery of nucleic acids with L2K resulted in high cellular concentrations of the nucleic acid cargo as well as a high fold-change in reporter activation experiments, but reactions likely occurred in endosomes during nucleic acid uptake. Delivering reactants sequentially with lipofectamine ensured that they only encounter one another in the cell. Not surprisingly, this approach resulted in a lower degree of co-localization and, consequently, lower fold-changes of activation. Sequential SLO delivery ensured that the nucleic acids are delivered directly into the cytosol, thus facilitat-

ing interactions with endogenous mRNA. However, the resulting nucleic acid concentrations in the cell are considerably lower than those achievable with lipid based delivery. To experimentally observe reporter activation, we thus co-localized multiple reporter gates on a native mRNA scaffold. Although experimentally more challenging, this approach provides the highest degree of control over the subcellular distribution of the delivered nucleic acids.

By characterizing elementary reporters and logic gates in cells and, importantly, demonstrating that such devices can be controllably interfaced with cellular mRNA, we have laid a foundation for the rational design of strand exchange-based control circuitry for mammalian cells. We believe that the unique advantages of strand exchange systems e.g. the small nucleotide footprint, the feasibility of rational design at the sequence level and the ease with which additional, orthogonal instances of any device can be created simply by changing the sequence will make them an important complement to existing approaches in RNA-based synthetic biology.

3.11 Future work.

As we have made a first step towards porting DNA circuitry into mammalian cells, DNA nanotechnology is still at an early stage for in cell use. Here I discuss two key challenges for advancing the field and suggest a few potential routes to overcoming them.

3.11.1 Developing a reporter system with single-molecular sensitivity.

The detection of specific cellular nucleic acids has important applications on both basic research and biomedicine, but the low copy number of many RNA species makes the detection difficult. Developing a reporter system with single-molecule sensitivity may provide a solution. For example, engineering mRNA with multiple tandem repeats allows many fluorescent probes to co-localize on a single mRNA and thus achieve single-molecule sensitivity[66]. Santamgelo *et al.* also developed a multiply labeled tetravalent RNA imaging probes (MTRIPs) in which multiple fluorophore labeled strands are attached to a streptavidin core. Hybridization of multiple MTRIPs to a single target can dramatically increase the signal to noise ratio, enabling the visualization of a single mRNA. Combined with other signal amplification systems such as catalytic amplifiers[41] and HCR amplifiers[46], it seems likely this challenge will be overcome.

3.11.2 Developing spatially organized circuits for the use in cells.

DNA strand displacement circuit components freely float in an aqueous environment, and wire-like connections are achieved by designing strands with complementary sequences. The operation of multi-stage circuits involves the interaction of several components, which requires all parts to freely diffuse in solution. However, since cells are highly structured and densely packed with macromolecules, circuit components often have a hard time encountering one another.

Instead of using freely diffusing components, we can make the various parts in close proximity and facilitate interactions by organizing them on DNA origami[27], a large-size

two- dimensional rectangular structure. Moreover, other groups have shown that large DNA nanostructures can enter living cells without a delivery agent[127], and DNA origami structures show great stability against degradation by nucleases[128]. Combining the use of DNA origami as a carrier with spatially organizing circuits will provide a new path for engineering multi-stage circuits inside cells.

3.12 Methods.

3.12.1 Sequence design

Sequences were designed using NUPACK, emphasis was placed on sequences that did not exhibit secondary structures, particularly in the toehold region. HPLC purified oligos were ordered from IDT with the indicated modifications.

3.12.2 Cell free kinetics assays.

All *in vitro* kinetics experiments were performed in Tris-acetate-EDTA buffer containing 12.5 mM Mg^{2+} (1x TAE/Mg^{2+}) at 37°C or PBS. Kinetics measurements were performed in a total 600 μl on a spectrofluorimeter (SPEX Fluorolog-3, Horiba) with 0.875 mL synthetic quartz cells (Starna catalog number 23-5.45-S0G-5). Excitation and emission values for different fluorophores were set as follows: ROX (588 nm/ 608 nm), Cy5 (648 nm/ 668 nm), TYE665 (645 nm/ 665 nm). The integration time was 10 sec for all experiments for every 60 sec time-point. A non-reactive 20 nucleotide poly-T carrier strand at the concentration of 1 μM was used in all reactions to reduce non-specifically binding to pipette tips.

3.12.3 Cell lines and cell culture.

CHO K1 cells were cultured in MEM/alpha media (Corning) + 10% FBS (Hyclone), 100 U/ml penicillin, and 100 μg/ml streptomycin (Invitrogen). HeLa cells and A549 cells were cultured in DMEM (Lonza) with 10% FBS, 100 U/ml penicillin, and 100 g/ml streptomycin (Invitrogen). For flow cytometry experiments, cells were plated at 50,000 cell/well in 24 well plates. For microscopy, cells were plated on glass coverslips one day prior to experiments.

3.12.4 Co-transfection.

When transfecting <150 nM of the indicated nucleic acid complex, 2 μl/well of Lipofectamine 2000 (L2K; Life Technologies), was used, while 3 μl was used when transfecting $\geq$150 nM.

In all cases, 1x refers to a 50 nM final concentration. Each complex was prepared with L2K independently and the resulting L2K/nucleic acid mixtures were then added to the cells together. Transfection were performed in pre-warmed Opti-MEM without phenol red. Transfections were allowed to proceed for 4 hours, following which cells were washed twice with 1x PBS. Finally, 500 μl of MEM/alpha media + 10% FBS was added and the cells were incubated for an additional 2 hours.

3.12.5 Sequential transfection.

The same ratios of transfection reagent to nucleic acid were used for co-transfection but but Lipofectamine RNAiMAX (Life Technologies) instead of L2K because of its lower toxicity. Transfection mixtures were prepared as above; the first transfection mixture was added and allowed to incubate with the cells for 2 hours. The cells were then washed twice with 1x PBS and the second transfection was performed. After proceeding for 2 hours, cells were washed twice again with 1x PBS and MEM/alpha media + 10% FBS was added. A final two hour incubation was performed before measurement.

3.12.6 Plasmid DNA transfection.

Plasmid DNA was transfected using Lipofectamine LTX (Life Technologies). 500 μg/well of plasmid DNA was used, along with 2 μl of Lipofectamine LTX. Transfections were allowed to proceed for at least 6 hours. Measurements were taken the following day.

3.12.7 SLO delivery.

mMTRIPs were sequentially delivered into A549 cells using reversible membrane permeabilization. Briefly, 4 U/ml Streptolysin O (SLO) (Sigma) were first reduced using 7.5mM Tris (2-carboxyethyl) phosphine (TCEP) (Pierce) for 1 hr at 37°C. Cells were rinsed using PBS (-Ca^{2+} -Mg^{2+}) (Thermo) and then incubated with delivery medium containing 0.4 U/ml SLO and probes at 15 nM in OptiMEM (Gibco) for 10 min at 37°C. The delivery medium

was then removed and replaced with DMEM for recovery for 1h. 1x or 3x molar excess of the appropriate input gate was subsequently delivered using SLO. Cells were fixed using 4% paraformaldehyde and nuclei were stained with DAPI and mounted for imaging.

3.12.8 siRNA sensor plasmids.

The plasmids used in this study were created using the Gibson cloning method. Each plasmid contains a pUC1 origin, ampicillin resistance marker, hygromycin marker (without a promoter), H2B-mCherry driven by the CMV promoter, and pEF-1α driving H2B-citrine with a modified version of the mus musculus Vamp3 3'UTR into which was inserted a target site for one or the other possible siRNA products. We independently created target plasmids for both possible mature siRNA that could result from the initial duplex, since either strand could potentially be inserted into the RISC complex. Maps are available as GenBank files online.

3.12.9 mMTRIP probes.

Monovalent MTRIPs (mMTRIPS) consist of a Neutravidin core bound to an mRNA-targeting ligand via an aromatic hydrazine and aldehyde linkage (Hynic-4FB, Solulink) and to four biotinlated and fluorescently labeled reporter oligonucleotides. The mRNA-targeting ligand containing a 5' thiol modification (Biosearch Technologies) are first conjugated to a 30x molar excess of maleimide HyNiC crosslinker (MHPH, Solulink). NeutrAvidin (Sigma) is labeled with 4FB groups (S-4FB, Solulink) to achieve a ratio of 2 4FB groups per molecule, as quantified by colorimetric reaction with 2-hydrazinopyridine (Solulink) by UV-Vis (Abs=350). The oligonucleotides and NeutrAvidin are separately filtered and buffer exchanged then combined according to the manufacturer protocol. The concentration of the resulting modified NeutrAvidin is quantified using a BCA Protein Assay Kit (Pierce). The reporter oligonucleotides contain a 5'-biotin modification and one dT-C6-NH2 modification for conjugation with Cy3B-NHS ester (GE Healthcare) using manufacturer protocols. Free dye was removed using 3kD Nanosep spin columns (Pall Corp.). The purified ligands were resuspended in

1xPBS and mixed at a 5:1 molar ratio with the modified NeutrAvidin for 1 hour at RT. Free ligands were removed using 30 kD Nanosep spin columns.

3.12.10 Confocal microscopy of fixed cells.

In all imaging experiments where the gates were introduced using the Lipofectamine 2000 transfection reagent, cells were imaged using a laser scanning confocal microscope Zeiss LSM 700 with a 40x objective 1.3NA. Laser combinations were as follows: DAPI/TYE665 = 405/639 for two-color simultaneous acquisition, DAPI/Alexafluor 488/TYE665 = 405/488/639 for co-localization measurements. In all cases we used the ZEN 2012 best compromise feature that optimizes filters for maximum acquisition speed with minimum crosstalk. mMTRIP AND gates were imaged using a Zeiss Axiovert 200 M microscope with a Perkin Elmer Ultraview spinning disk confocal using a 63xNA 1.4 Plan Apochromat primary objective and a Hammamatzu Flash 4.0v2 camera.

3.12.11 Fluorescence imaging of fixed cells.

Images for 4-way experiments were taken on a Zeiss Axiovert 200M epifluorescence microscope with a 63x NA 1.4 Plan Apochromat primary objective and a Hammamatsu ORCA-ER AG camera. The Chroma Sedat ET filter set was used. All imaging experiments were performed using the Volocity acquisition software (Improvision, PerkinElmer). Image z-stacks were acquired in 200nm steps. All images were deconvolved using the iterative algorithms in Volocity. Images are extended views, which compress all of the imaging planes in the z direction into one plane. The mean sum of intensity was used to determine changes in fluorescence intensity in all experimental conditions.

3.12.12 Flow cytometry

To detach cells, each well was washed once with 500 μl 1x PBS then added 100 μl of 0.25% trypsin + EDTA were added. Cells were incubated for 5 min and the trypsin quenched

using 200 μl 1x PBS + 2% FBS. Cells were triturated several times before being run on an Accuri C6 cytometer (BioRad). Gates were drawn using FSC-H vs. FSC-A and were kept the same throughout the experiments.

3.12.13 Flow cytometry data processing

For all flow cytometry experiments we made use of a BD Accuri C6 flow cytometer. FL1-A measures excitation from a 488 nm laser through a 533/30 filter. We used this to quantify the signal from Citrine (fluorescent protein) and ATTO488N (fluorophore label). FL3-A measures excitation from the 488 nm laser through a 610/20 filter (in our experience, this yields qualitatively similar results to using FL2-A, which uses the same laser, but a 585/40 filter). We used this filter/laser combination to measure mCherry (fluorescent protein) and/or the TYE563 fluorophore. FL4-A pairs a 640 nm laser with a 675/25 filter. This was used to measure the signal from the TYE665 fluorophore.

A gate based on forward scattered-height (FSC-H) vs. forward scattered-area (FSC-A) was used to capture a population of a somewhat uniform size. For CHO K1 cells, the gate was generally drawn between 1.5-4$*10^6$ FSC-H and 2-3.8$*10^6$ FSC-A (Fig. 3.16a). This particular gate was used in all experiments in which we used CHO K1 cells. When analyzing HeLa cells, a different gate was drawn in the same manner.

Compensation for signal spillover is performed as follows. We measured cells delivered with single fluorophore-labeled 2'Ome RNA complexes for each channel, and estimate the spillover. Specifically, we used (1) the ATTO488N fluorophore to measure the FL1 spillover (Fig. 3.16b), (2) the TYE563 fluorophore to measure the FL3 spillover (Fig. 3.16c), and (3) the TYE665 fluorophore to measure the FL4 spillover (Fig. 3.16d). Some degree of spillover was observed for all channels, though to a much greater degree between green and red (FL1 and FL3); signal spillover was then fitted to estimate the signal crossover slopes. Once the slope of spillover is known, a spillover matrix (SM) can be made as follows:

SM $= \begin{pmatrix} 1 & s_{12} & s_{13} \\ s_{21} & 1 & s_{23} \\ s_{31} & s_{32} & 1 \end{pmatrix}$, where s_{12} is the slope of spillover from FL3 into FL1, s_{13} is the slope of the spillover from FL4 into FL1, s_{21} is the slope of spillover from FL1 into FL3, etc..

3.12.14 Cell viability staining

Cell viability staining was performed following the manufacturer's instructions. Briefly, following the indicated treatments, cells in 24 well plates were treated with 100 μl of 0.25% trypsin at 37°C for ∼5 min to dissociate them from the plate surface. The trypsin was quenched with 200 μl 1x PBS + 2% FBS + 0.2 μl/ml SYTOX Green dead cell stain (Life Technologies Cat# S34860). The plate was immediately protected from light and cooled to 4°C and incubated for 1 hour. Following the incubation the fluorescence was quantified using a BD Accuri C6 Flow cytometer.

3.12.15 Gating of CHO K1 cells transfected with siRNA sensor plasmid

CHO K1 cells were transfected with the sensor plasmid followed by the siRNA-activation gates. Cells were fixed and analyzed via flow cytometry. Since not all of the cells in the population were successfully transfected with the plasmid, we drew a threshold gate to largely exclude the un-transfected population (vertical blue line in Fig. 3.17b). Only cells above this threshold were counted in our analysis. Scatter plots of the cell data used to calculate the values reported in the main text appear in Fig. 3.17c.

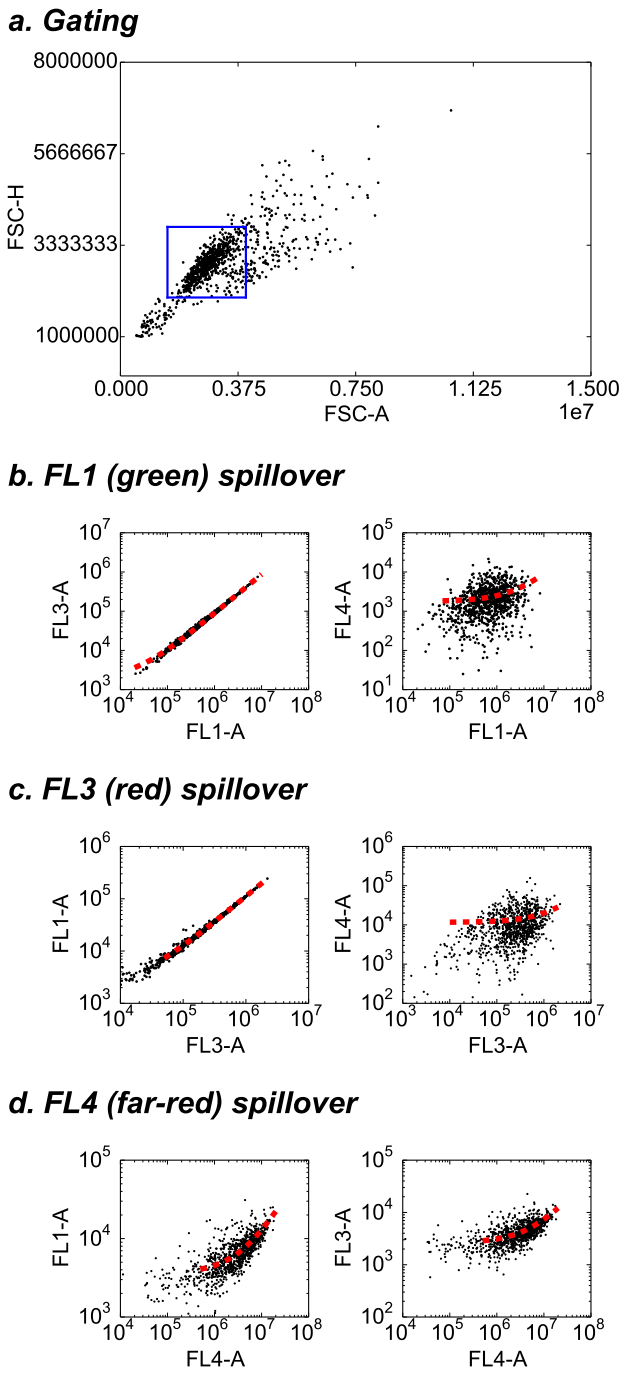


Figure 3.16: Flow cytometry data processing. **a,** A scatter plot showing forward scatter area (FSC-A) versus forward scatter height (FSC-H) was generated, and cells were gated between 1.5-4$*10^6$ FSC-H and 2-3.8$*10^6$ FSC-A. b-d. Cells delivered with fluorophore labeled 2'Ome RNA complexes to estimate the signal crossover. b. Spillover from FL1 into FL3 and FL4. c. Spillover from FL1 into FL3 and FL4. d. Spillover from FL1 into FL3 and FL4. The fit to the spillover is shown as red dashed lines.

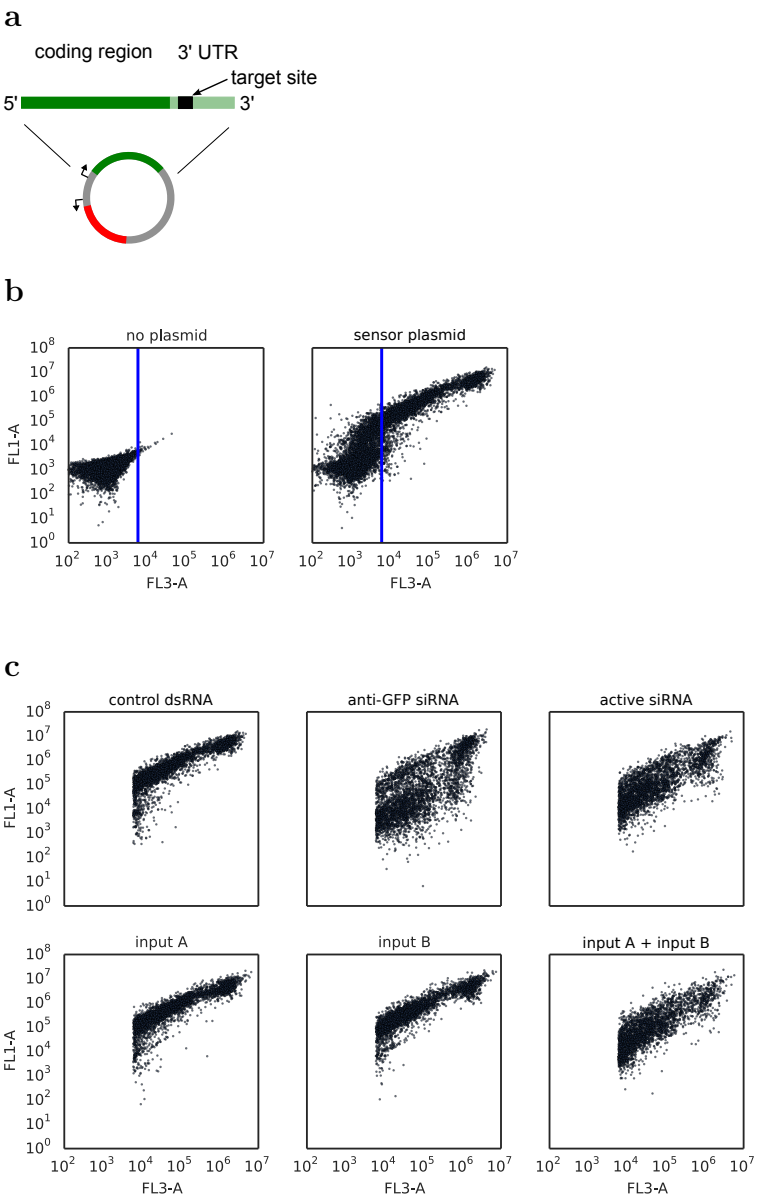


Figure 3.17: Threshold gating of cells transfected with the siRNA sensor plasmid. **a,** Sensor plasmid. The target site of siRNA was inserted to the 3'UTR of mCitrine. The mCherry signal is used as a reference signal to adjust cell-to-cell variation in plasmid delivery. **b,** Data from flow cytometry measurements of CHO K1 cells either without the sensor plasmid or transfected with sensor plasmid; both sets were subsequently transfected with a double-stranded control RNA as well. A vertical threshold gate was drawn to exclude cells that with no measurable signal from the sensor plasmid. **c,** Gated data from cells treated with the sensor plasmid and then transfected with the indicated complexes. Samples with only one indicated gate were also transfected with a control RNA. Green and yellow signals (H2B-mCitrine) are measured in FL1-A, while FL3-A is equivalent to the red (H2B-mCherry) signal. The mCitrine signal from each cell was normalized to its mCherry signal. These values were then averaged for each sample. At least two samples were collected for each condition during each experiment, and at least two experiments were performed.

3.13 Acknowledgements.

We thank David Soloveichik, Erik Winfree, and David Yu Zhang for their help in designing the logic AND gate. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. W911NF3113230068 to GS and PJS. Additional support for was provided by NSF CAREER (PJS) and NIH GM094198 (PJS).

Table 3.1: Domain sequences of the 4-way strand exchange systems. All sequences start from the 5' end. The number in parenthesis followed by an alphabetic letter indicates the length of a long domain. For example, "a(22)" indicates the length of domain "a" is 22-bp.)

Domain	**Sequence**	**Length (nt)**
1	GGGAAT	6
2	TCTTAC	6
3	TGATGA	6
4	AACTAC	6
5	AATGAC	6
6	TTACCT	6
a(22)	GTAGGAGTGGAGGTGATGATGA	22
a(16)	GTAGGAGTGGAGGTGA	16
a(10)	GTAGGAGTGG	10
b(16)	CAACACACACACACCC	16
b(10)	CAACACACAC	10

Table 3.2: Strand sequences of different covalent modifications for our 4-way strand exchange system in Fig. 3.4.

Strand	**Sequence**
<2* a>-TYE665 (DNA)	GTAAGAGTAGGAGTGGAGGTGA/3TYE665/
RQ-<a* 1*> (DNA)	/5IAbRQ/TCACCTCCACTCCTACATTCCC
<a* 2> (DNA)	TCACCTCCACTCCTACATTCCCACTCTTAC
<1 a> (DNA)	GGGAATGTAGGAGTGGAGGTGA
<2* a>-TYE665 (PS DNA)	G*T*A*A*G*A*G*TAGGAGTG*G*A*G*G*T*G*A /3TYE665/
RQ-<a* 1*> (PS DNA)	/5IAbRQ/T*C*A*C*C*T*C*CACTCCTA*C*A*T*T*C*C*C
<a* 2> (PS DNA)	T*C*A*C*C*T*C*CACTCCTA*C*T*C*T*T*A*C
<1 a> (PS DNA)	G*G*G*A*A*T*G*TAGGAGTG*G*A*G*G*T*G*A
<2* a>-TYE665 (2'OMe RNA)	mGmUmAmAmGmAmGmUmAmGmGmAmGmUmGmGmAmGmGmUmGmA /3TYE665/

RQ-<a* 1*> (2'OMe RNA)	/5IAbRQ/ mUmCmAmCmCmUmCmCmAmCmUmCmC-mUmAmCmAmUmUmCmCmC
<a* 2> (2'OMe RNA)	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mUmCmUmUmAmC
<1 a> (2'OMe RNA)	mGmGmGmAmAmUmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA
<2* a>-TYE665 (PS 2'OMe RNA)	mG*mU*mA*mA*mG*mA*mG*mUmAmGmGmAmGmU mG*mG*mA*mG*mG*mU*mG*mA/3TYE665/
RQ-<a* 1*> (PS 2'OMe RNA)	/5IAbRQ/ mU*mC*mA*mC*mC*mU*mC*mCmAmCmU mCmCmUmA*mC*mA*mU*mU*mC*mC*mC
<a* 2> (PS 2'OMe RNA)	mU*mC*mA*mC*mC*mU*mC*mCmAmCmUmCmCmU mA*mC*mU*mC*mU*mU*mA*mC
<1 a> (PS 2'OMe RNA)	mG*mG*mG*mA*mA*mU*mG*mUmAmGmGmAmGmU mG*mG*mA*mG*mG*mU*mG*mA

Table 3.3: Strand sequences of three-color 4-way strand exchange controls. Strands are all 2'OMe RNA.

Strand	**Sequence**	**Length (nt)**
<a* 2>-ATTO488	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mUmCmUmUmAmC /3ATTO488N/	22
<1 a>	mGmGmGmAmAmUmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	22
<2* a>-TYE665	mGmUmAmAmGmAmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA /3TYE665/	22
RQ-<a* 1*>-TYE563	/5IAbRQ/ mUmCmAmCmCmUmCmCmAmCmUmCmC-mUmAmCmAmUmUm CmCmC/3TYE563/	22
<a* 5>-ATTO488	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mAmAmUmGmAmC /3ATTO488N/	22
<4 a>	mAmAmCmUmAmCmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	22
<b* 2>-ATTO488	mGmGmGmUmGmUmGmUmGmUmGmUmGmUmUmG-mUmCmUmUmAmC /3ATTO488N/	22
<1 b>	mGmGmGmAmAmUmCmAmAmCmAmCmAmCmAmC-mAmCmAmCmCmC	22

Table 3.4: Strand sequences for our AND logic gate implemented using 2'OMe RNA. The length of the domain "a" here is 10 bp. The corresponding DNA domain sequences are in Table 3.1.)

Domain	**Sequence**	**Length (nt)**
<3* b 2* a>-TYE665	mUmCmAmUmCmAmCmAmAmCmAmCmAmCmAm-CmAmCmAmCmCmCmGmUmAmAmGmAmGmU-mAmGmGmAmGmUmGmGmAmGmGmUmGmA /3TYE665/	32
RQ-<a* 1* b* 4*>	/5IAbRQ/ mUmCmAmCmCmUmCmCmAmCmUmCmC-mUmAmCmAmUmUmCmCmCmGmGmGmUmGmUmG-mUmGmUmGmUmGmUmUmGmGmUmAmGmUmU	32
<a* 2>	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mUmCmUmUmAmC	16
<1 a>	mGmGmGmAmAmUmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	16
<a* 5>	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mAmAmUmGmAmC	16
<6 a>	mUmUmAmCmCmUmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	16
<4 b>	mAmAmCmUmAmCmCmAmAmCmAmCmAmCmAmC-mAmCmAmCmCmC	16
<b* 3>	mGmGmGmUmGmUmGmUmGmUmGmUmGmUmUmG-mUmGmAmUmGmA	16
<6 b>	mUmUmAmCmCmUmCmAmAmCmAmCmAmCmAmC-mAmCmAmCmCmC	16

<b* 5>	mGmGmGmGmUmGmUmGmUmGmUmGmUmGmUmUmG-mAmAmUmGmAmC	16

Table 3.5: Strand sequences for our OR hybrid-2. The length of both domains "a" and "b" are 16-bp. The corresponding DNA domain sequences are in Table 3.1.

Domain	**Sequence**	**Length (nt)**
<3* b a* 1*>	mUmCmAmUmCmAmCmAmAmCmAmCmAmCmAm-CmAmCmAmCmCmCmUmCmAmCmCmUmCmC-mAmCmUmCmCmCmUmAmCmAmUmUmCmCmC	44
<2* a>-Cy5	mGmUmAmAmGmAmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA /3Cy5Sp/	22
RQ-<b* 4*>	/5IAbRQ/ mGmGmGmUmGmUmGmUmGmUmG-mUmGmUmUmGmGmUmAmGmUmU	22
<a* 2>	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mUmCmUmUmAmC	22
<1 a>	mGmGmGmAmAmUmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	22
<a* 5>	mUmCmAmCmCmUmCmCmAmCmUmCmCmUmAmC-mAmAmUmGmAmC	22
<4 a>	mAmAmCmUmAmCmGmUmAmGmGmAmGmUmGmG-mAmGmGmUmGmA	22
<b* 3> (PS 2'OMe RNA)	mG*mG*mG*mU*mG*mU*mG*mUmGmUmGmUmGmU mU*mG*mU*mG*mA*mU*mG*mA	22
<4 b>	mAmAmCmUmAmCmCmAmAmCmAmCmAmCmAmC-mAmCmAmCmCmC	22
<6 b>	mUmUmAmCmCmUmCmAmAmCmAmCmAmCmAmC-mAmCmAmCmCmC	22
<b* 5> (PS 2'OMe RNA)	mG*mG*mG*mU*mG*mU*mG*mUmGmUmGmUmGmU mU*mG*mA*mA*mU*mG*mA*mC	22

Table 3.6: Strand sequences for mTRIP gates. Multivalent probes are labeled with a 5'biotin; monovalent probes are labeled with a 5'thiol.

RNA target	**Ligand sequence**	**Location within the mRNA sequence**
c-Myc 3'UTR A 1	5'-TXTTTT**GAAXCGUXUUCCXUACUXUUC**-3'	1892-1912
c-Myc 3'UTR A 2	5'-TXTTTT**CAAGXUCAXAGGXGAUXGCUC**-3'	1932-1952
c-Myc 3'UTR A 3	5'-TXTTTT**GGXUGXGAGGXUGCAUXUG**-3'	1968-1986
Poly(A)	5'-TXTTTT**UXUUUUUUXUUUUUUXUU**-3'	
β-actin 3'UTR 1	5'-TXTTTT**XUCCUGXAACAAXGCAUCXC**-3'	1479-1461
β-actin 3'UTR 2	5'-TXTTTTX**AAGCCCXGGCXGCCXCCA**-3'	1748-1730
β-actin 3'UTR 3	5'-TXTTTT**GXCACCUXCACCGUXCCAG**-3'	1377-1359
β-actin 1	5'-UXTTTX**AXAGCACAGCCXGGAXA**-3'	494-478
β-actin 2	5'-TTTTTTX**AUUXCCCGCXCGGCCGXG**-3'	696-679
4-way reporter	5'-**GUAAGAGUAGGAGUGGAGGUGA**-X-TTTTTT-biotin-3'	
AND reporter	5'-**UCAUCACAACACACACACACCCG UAAGAGUAGGAGUGGAGGUGA**-X-TTTTTT-biotin-3'	

	Boldface: 2'-O-Methyl RNA; X: dT-C6-NH2; all others are DNA; underline: binding region.	

Chapter 4

APPENDIX

4.1 The convergence of the kinetics of the strand displacement-level model to the target CRN

We desire that the DNA realization of a formal CRN implemented according to our design schema behaves quantitatively close to the target CRN. This section substantiates this expectation of kinetic equivalence with a theoretical argument based on the strand displacement-level model. While we cannot experimentally build and test all CRNs, this argument supports the generality of our construction even for untested CRNs.

We show that the strand displacement-level model converges to the desired CRN kinetics in the limit of high concentration of gates and auxiliary species relative to the concentrations of the signal species ("CRN regime"). (The strand displacement model describes each strand displacement event as a elementary reaction. This is the mechanistic model used throughout this paper. We start with a single reaction $A + B \rightarrow C$, and then generalize to arbitrary systems of bimolecular reactions in the following section. A similar argument was developed in [34], but for a different strand displacement implementation of CRNs. The argument presented here is an informal one (for a more rigorous argument, a singular perturbation analysis could be used as in [34].) The convergence argument also allows us to derive the effective rate constants for the formal CRN reactions as functions of strand displacement-level rate constants and concentrations, providing a systematic way to program rates.

4.1.1 Theory of convergence for $A + B \to C$

We first outline an argument that our implementation of the formal reaction $A + B \to C$ satisfies the ideal bimolecular reaction rate law

$$\frac{d[C]}{dt} = -\frac{d[A]}{dt} = -\frac{d[B]}{dt} = [A][B]k.$$

The appropriate regime is one with high relative concentration of gates and auxiliary species. The bimolecular rate constant k is given by equation 4.8 below. This supports the kinetic correctness of our implementation of the formal reaction, and further equation 4.8 can be used to design the kinetic rate constant of the target reaction. In the next section, we provide numerical simulations illustrating this convergence, and conclude with experimental evidence corroborating our derivation.

We make a number of assumptions, the most important of which is that the gates, auxiliary, and backward auxiliary species (Join_{AB}, Fork_C, $\langle tr\ r\rangle$, $\langle i\ tc\rangle$, $\langle a\ tb\rangle$, $\langle b\ tr\rangle$, $\langle c\ tr\rangle$) are in large excess compared to the concentrations of the signal species. These amounts are assumed to be large enough that they stay effectively constant during the time-course of the experiment, and thus their instantaneous concentrations $[\cdot]$ are the same as their initial concentrations $[\cdot]_0$. Further, we suppose the excess is large enough that a strand displacement reaction involving one of these high concentration species proceeds much faster than any strand displacement reaction between other species. Per our strand displacement-level model, each strand displacement step is a bimolecular elementary reaction $G + \langle S\rangle \to G' + \langle S'\rangle$ where G is the original gate, $\langle S\rangle$ is the displacing strand, $\langle S'\rangle$ is the displaced strand, and G' is the resulting gate. The strand displacement reaction is reversible or irreversible depending on whether or not it leaves a toehold. For generality, we allow each strand displacement rate constant to be different.

For this analysis we need to be careful about distinguishing formal species A, B, C from their corresponding signal strands, which we'll write as $\langle A\rangle, \langle B\rangle, \langle C\rangle$. We'll use the notation Join_{AB}-i to mean the Join_{AB} gate after i strand displacements (i.e. with the $i+1$st toehold from the left open). Similarly, for the Fork_C gate, we write Fork_C-i to mean the gate after

i strand displacements (i.e. with the $i+1$st toehold from the right open, except for Fork$_C$-3 which is fully double-stranded). To aid classifying which reactions are relatively fast and which are relatively slow, we box the reactants present in high concentration. We assume that the experiment does not proceed long enough to change these designations. Then the Join$_{AB}$ strand displacements are modeled by the following reactions:

$$\boxed{\text{Join}_{AB}} + \langle A\rangle \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \text{Join}_{AB}\text{-}1 + \boxed{\langle a\ tb\rangle} \tag{4.1}$$

$$\text{Join}_{AB}\text{-}1 + \langle B\rangle \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \text{Join}_{AB}\text{-}2 + \boxed{\langle b\ tr\rangle} \tag{4.2}$$

$$\text{Join}_{AB}\text{-}2 + \boxed{\langle tr\ r\rangle} \underset{k_6}{\overset{k_5}{\rightleftharpoons}} \text{Join}_{AB}\text{-}3 + \langle r\ tq\rangle \tag{4.3}$$

and the Fork$_C$ strand displacements by the following reactions:

$$\boxed{\text{Fork}_C} + \langle r\ tq\rangle \underset{k_8}{\overset{k_7}{\rightleftharpoons}} \text{Fork}_C\text{-}1 + \boxed{\langle tr\ r\rangle} \tag{4.4}$$

$$\text{Fork}_C\text{-}1 + \boxed{\langle c\ tr\rangle} \underset{k_{10}}{\overset{k_9}{\rightleftharpoons}} \text{Fork}_C\text{-}2 + \langle C\rangle \tag{4.5}$$

$$\text{Fork}_C\text{-}2 + \boxed{\langle i\ tc\rangle} \xrightarrow{k_{11}} \text{Fork}_C\text{-}3 + \langle i\rangle. \tag{4.6}$$

The slowest reaction along the forward path is reaction 4.2 because it is the only forward reaction in which both reactants are present in small amounts (unboxed). The rate of this reaction is [Join$_{AB}$-1][$\langle B\rangle$]k_3. To see how the rate of producing $\langle C\rangle$ is dependent on the rate of this rate-limiting step consider the following Markov chain analysis on a single Join$_{AB}$ gate molecule. After an instance of forward reaction 4.2 occurs, there is a $[\langle tr\ r\rangle]_0 k_5/([\langle b\ tr\rangle]_0 k_4 + [\langle tr\ r\rangle]_0 k_5)$ chance that reaction 4.3 occurs next releasing $\langle r\ tq\rangle$ before reaction 4.2 reverses. Thus the rate of producing $\langle r\ tq\rangle$ is roughly

$$[\text{Join}_{AB}\text{-}1][\langle B\rangle]k_3 \frac{[\langle tr\ r\rangle]_0 k_5}{[\langle b\ tr\rangle]_0 k_4 + [\langle tr\ r\rangle]_0 k_5}. \tag{4.7}$$

After $\langle r\ tq\rangle$ is released, the slow reverse of reaction 4.3 (neither reactant is in high concentration) constitutes a barrier to undoing the preceeding progress. Thus we can suppose that once $\langle r\ tq\rangle$ is released, the strand displacement cascade 4.4-4.6 will proceed to the end. If the rate-limiting step is significantly slower than any step involving a high concentration (boxed)

species, then we can ignore the delays due to the fast steps. In this regime expression 4.7 approximates the rate of producing $\langle C \rangle$.

Finally, we establish the relationship between [Join$_{AB}$-1] and $[A]$. Assume that reaction 4.1, which is fast in both directions, reaches quasi-equilibrium on a time scale faster than that of the rate limiting step (forward reaction 4.2). We can think of Join$_{AB}$-1 and free signal strand $\langle A \rangle$ as two forms of the formal species A, with $[A] = [\langle A \rangle] + [\text{Join}_{AB}\text{-1}]$. This is the amount of A that would be recorded when a large amount of reporter for $\langle A \rangle$ is added: Join$_{AB}$-1 reverts to $\langle A \rangle$ and reacts with the reporter on a faster time scale than reaction 4.2 proceeds forward. Then at quasi-equilibrium, the amount of A in the system will be divided between these two forms with $[\text{Join}_{AB}\text{-1}] = [A][\text{Join}_{AB}]_0 k_1 / ([\text{Join}_{AB}]_0 k_1 + [\langle a\ tb \rangle]_0 k_2)$. No similar quasi-equilibrium exists for B and thus $[\langle B \rangle] = [B]$. Plugging this into expression 4.7 yields

$$[A][B]k \text{ where } k = k_3 \frac{[\text{Join}_{AB}]_0 k_1}{[\text{Join}_{AB}]_0 k_1 + [\langle a\ tb \rangle]_0 k_2} \cdot \frac{[\langle tr\ r \rangle]_0 k_5}{[\langle b\ tr \rangle]_0 k_4 + [\langle tr\ r \rangle]_0 k_5}. \tag{4.8}$$

Equation 4.8 is our final estimate of the effective rate of producing $\langle C \rangle$, and consuming A (in either form) and $\langle B \rangle$. k is the effective bimolecular rate constant of the resulting formal bimolecular reaction $A + B \xrightarrow{k} C$. Note that, in the simplest case, if using equal amounts of all high concentration species, and $k_1 = k_2$, $k_4 = k_5$, then $k = k_3/4$ — that is the overall kinetics is that of the second forward strand displacement step slowed down by a factor of 4.

If our reaction $A + B \to C$ were a part of a larger system of coupled reactions, such that A or B participated as a first reactant somewhere else, then at quasi-equilibrium some amount of A and B would be sequestered in other Join gates. (See the "buffering effect" discussed in [34].) We generalize the above argument to an arbitrary bimolecular CRN in section 4.1.3.

4.1.2 Simulation of convergence for $A + B \to C$, and comparison to experimentally collected kinetics

4.1a visually demonstrates the convergence of the strand displacement-level model (ie reactions 4.1–4.6) to the bimolecular model with increasing concentration of the boxed (high

concentration) species. The most obvious deviation from ideal bimolecular kinetics at lower concentrations of the boxed species is the initial behavior of the numerical derivative, which obtains a maximum value only after a delay. There are at least two factors contributing to this delay. First, before the pseudo-equilibrium is established between $\langle A \rangle$ and Join$_{AB}$-1, the rate of the rate-limiting step is slower than expected. Second, our Markov chain analysis ignored the delays due to strand displacement steps other than the second strand displacement of the Join$_{AB}$ gate (the rate-limiting step). This further contributes to the observed delay of driving up the production of C. These delays decrease with larger initial concentrations of the boxed species because the pseudo-equilibrium is established faster and the non-rate limiting steps are sped up. The other factor contributing to the deviation from ideal bimolecular kinetics at lower concentration of the boxed species is that the concentrations of Join$_{AB}$, $\langle a\ tb \rangle$, $\langle b\ tr \rangle$ all change during the course of the simulation, while our analysis assumes they are constant. Again this approximation becomes more valid with higher concentrations of these species.

Lastly, we turn to the experimental data itself and ask how well does it corresponds to our bimolecular model and expression 4.8. We used the strand displacement level rate constants k_1–k_5 obtained as discussed in Section S7 in [35]. Fig. 4.1b shows that expression 4.8 agrees well with the bimolecular rate constant fit to experimental data for concentrations of $\langle a\ tb \rangle$, $\langle b\ tr \rangle$ over 1x. At smaller concentrations, the assumptions of our derivation are significantly violated — particularly that these concentrations stay constant throughout the experiment. Further, 4.1d shows the behavior of the numerical derivative of the production of C obtained from experimental data. As predicted from the strand displacement level model, the numerical derivative begins to agree with the fitted bimolecular rate law after an initial delay.

4.1.3 Theory of convergence for arbitrary bimolecular CRNs

In this section we generalize the convergence from Section 4.1.1, which was confined to a single reaction, to arbitrary bimolecular CRNs.

Consider the general form of a bimolecular reaction

$$\alpha + \beta \xrightarrow{k} \gamma + \cdots + \delta \tag{4.9}$$

where $\alpha, \beta, \gamma, \delta$ index over species (the same species can appear in multiple places.) Analogous to the $A + B \to C$ case discussed above, the Join gate strand displacement reactions can be written in general as follows:

$$\boxed{\text{Join}_{\alpha\beta}} + \langle\alpha\rangle \underset{k_{\alpha\beta2}}{\overset{k_{\alpha\beta1}}{\rightleftharpoons}} \text{Join}_{\alpha\beta}\text{-}1 + \boxed{\langle\text{back}_{\alpha}\rangle} \tag{4.10}$$

$$\text{Join}_{\alpha\beta}\text{-}1 + \langle\beta\rangle \underset{k_{\alpha\beta4}}{\overset{k_{\alpha\beta3}}{\rightleftharpoons}} \text{Join}_{\alpha\beta}\text{-}2 + \boxed{\langle\text{back}_{\beta}\rangle} \tag{4.11}$$

$$\text{Join}_{\alpha\beta}\text{-}2 + \boxed{\langle\text{aux}_{\alpha\beta}\rangle} \underset{k_{\alpha\beta6}}{\overset{k_{\alpha\beta5}}{\rightleftharpoons}} \text{Join}_{\alpha\beta}\text{-}3 + \langle\text{trans}_{\alpha\beta}\rangle \tag{4.12}$$

where the "trans" strand is then input to the Fork gate. Let $j_{\alpha\beta} = [\text{Join}_{\alpha\beta}]_0$, $a_{\alpha\beta} = [\text{aux}_{\alpha\beta}]_0$, $b_{\alpha} = [\text{back}_{\alpha}]_0$, be the initial concentrations of the gates, auxiliary, and backward auxiliary strands, respectively. As before we assume that these quantities are in large excess and do not vary over the time course of the experiment. Then the same Markov chain analysis as in the single reaction case yields that the rate of the production of $\langle\text{trans}_{\alpha\beta}\rangle$ (analog of equation 4.7) is

$$[\text{Join}_{\alpha\beta}\text{-}1]\,[\langle\beta\rangle]\,k_{\alpha\beta3}\,\frac{a_{\alpha\beta}\,k_{\alpha\beta5}}{b_{\beta}\,k_{\alpha\beta4} + a_{\alpha\beta}\,k_{\alpha\beta5}}. \tag{4.13}$$

In a regime with large concentrations of the Fork gate and auxiliary species, any trans strand produced is quickly converted to reaction products $\gamma + \cdots + \delta$, and therefore the rate of the production of the trans strand is a valid the measure of the overall reaction rate.

Formal species α is split up in a pseudo-equilibrium between $\langle\alpha\rangle$, $[\text{Join}_{\alpha\beta}\text{-}1]$, and the corresponding Join-1 of other formal reactions with α as the first reactant. Similarly, formal species β is split up in a pseudo-equilibrium between $\langle\beta\rangle$ and Join-1 of other formal reactions with β as the first reactant. So in equation 4.13, we can approximate $[\text{Join}_{\alpha\beta}\text{-}1]$ as a constant fraction of all of $[\alpha]$, and the free strands $[\langle\beta\rangle]$ as constant fraction of all of $[\beta]$ as follows. Let

$$g_{\alpha\beta} = \frac{j_{\alpha\beta}\,k_{\alpha\beta1}}{b_{\alpha}\,k_{\alpha\beta2}}.$$

Then

$$[\text{Join}_{\alpha\beta}\text{-}1] = [\alpha]\left(\frac{g_{\alpha\beta}}{1+\sum_{\beta'} g_{\alpha\beta'}}\right)$$

and

$$[\langle\beta\rangle] = [\beta]\left(\frac{1}{1+\sum_{\beta'} g_{\beta\beta'}}\right).$$

Note that the first sum is over all reactions that share the same first reactant as this reaction, and the second sum is over all reactions whose first reactant is the same as the second reactant of this reaction. Therefore, we can approximate the effective bimolecular rate constant of the implemented formal reaction (4.9) as:

$$k = k_{\alpha\beta 3}\left(\frac{g_{\alpha\beta}}{1+\sum_{\beta'} g_{\alpha\beta'}}\right)\left(\frac{1}{1+\sum_{\beta'} g_{\beta\beta'}}\right)\frac{a_{\alpha\beta}\, k_{\alpha\beta 5}}{b_\beta\, k_{\alpha\beta 4} + a_{\alpha\beta}\, k_{\alpha\beta 5}}. \tag{4.14}$$

4.1.4 Simulation of the convergence for the consensus network

4.2 shows the convergence of the strand displacement-level model to the formal three reaction CRN of the consensus network (Fig. 2.6b) with increasing concentration of the boxed (high concentration) species. The strand displacement-level model is described by reactions 4.10–4.12, and the analogous reactions of the Fork gate. The rate constants of the consensus three reaction CRN are obtain by expression 4.14 above. The scaling factor s is used to uniformly scale up the concentrations of the high concentration species. A larger scaling factor s corresponds to better agreement to the assumptions of our derivation ("CRN regime").

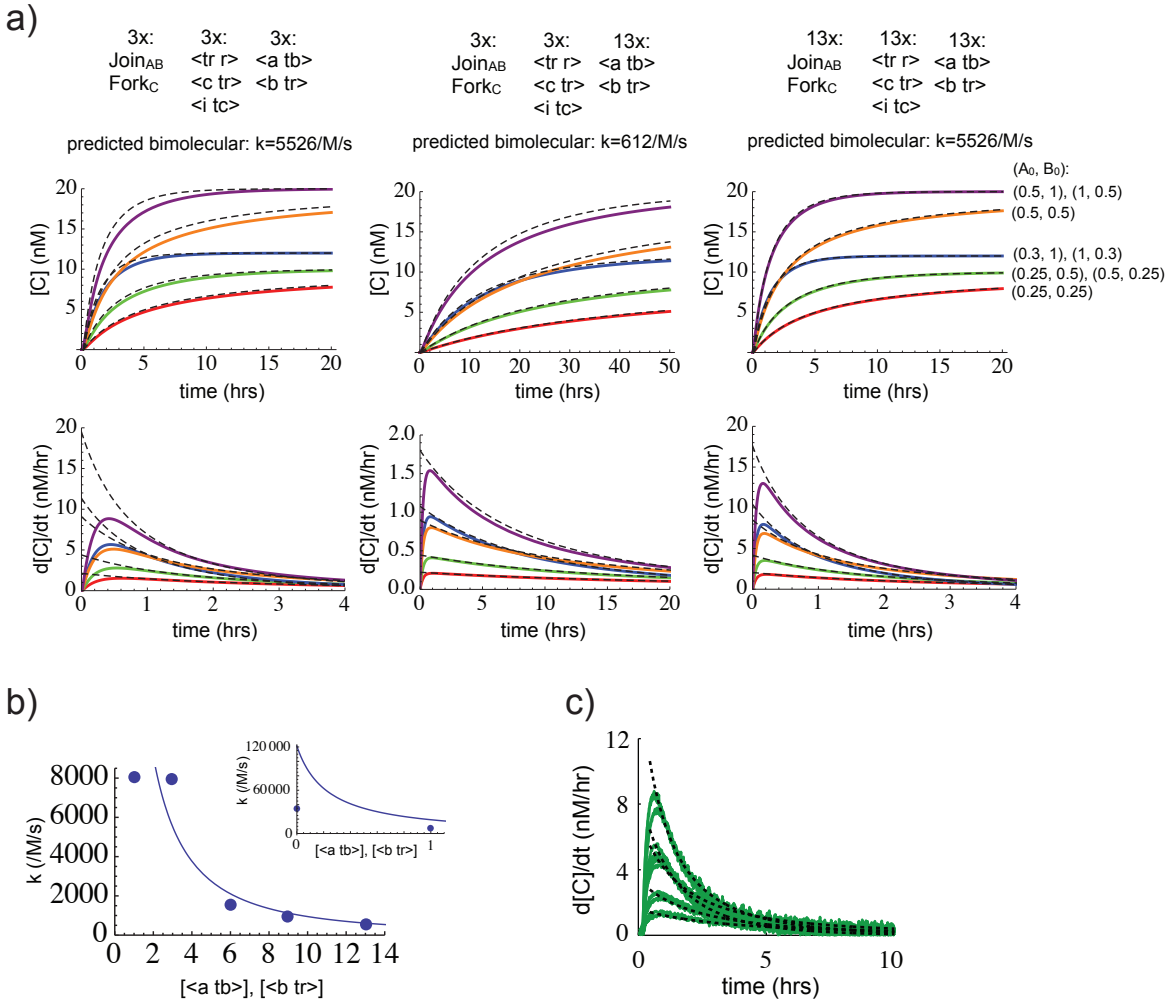


Figure 4.1: **a.** Convergence of the strand displacement-level simulation (i.e. reactions 4.1–4.6) to the bimolecular idealization. When the initial concentration of the boxed species is increased, the strand displacement-level simulation begins to be well-approximated by a simple bimolecular model $A + B \xrightarrow{k} C$ where k is given by expression 4.8. Dashed: bimolecular model. Solid: simulation of strand displacement reactions 4.1–4.6. **b.** The analytical prediction (expression 4.8) agrees well with the bimolecular rate constant fit to experimental data (shown in Fig. 2.5b) when the concentrations of $\langle a\ tb \rangle$, $\langle b\ tr \rangle$ are over 1x. The inset shows that the prediction deviates at 0x and 1x concentrations, because the system no longer satisfies the assumptions of our derivation. Dots: bimolecular rate constants fit to experimental data; solid line: value of expression 4.8. This figure is plotted on a log-scale in the main text (Fig. 2.5c). In both (a) and (b): 1x = 40 nM. Strand displacement-level rate constants are from section S7 in [35]. **c.** Comparison of experimentally measured numerical derivative and bimolecular rate law. The rate of the formation of product C is obtained by taking a time derivative of the data in Fig. 2.5b (3x case), and the fitted bimolecular model is plotted as black dashed traces. Note that the bimolecular approximation becomes valid after an initial transient during which individual reaction steps equilibrate.

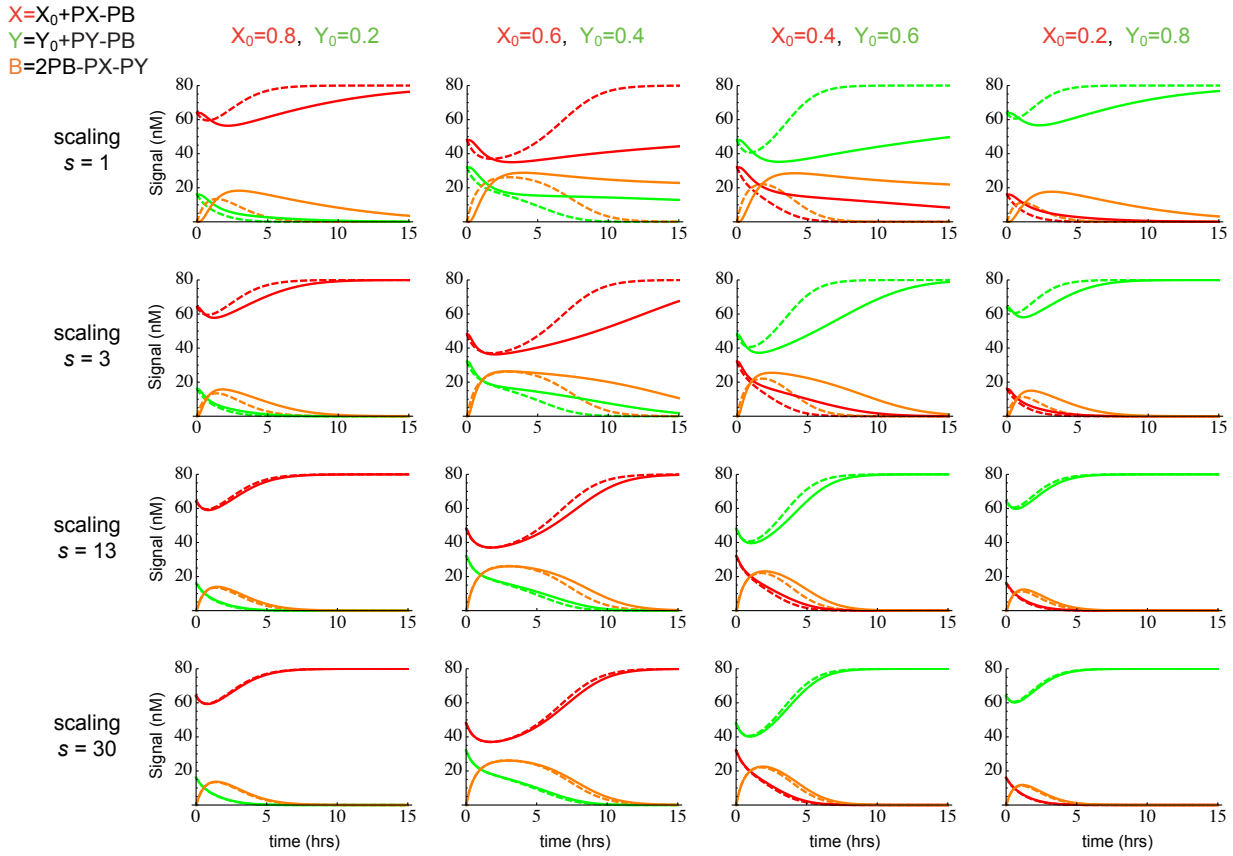


Figure 4.2: Convergence of the strand displacement-level simulation (reactions 4.10–4.12 and the analogous reactions of the Fork gate) to the formal CRN model for the consensus network. When the initial concentration of the boxed species is increased, the strand displacement-level simulation begins to be well-approximated by the three reaction CRN, where the rate constants of the three reactions are derived according to expression 4.14. Dashed: formal CRN model. Solid: strand displacement-level model. The concentrations used are as in the text (Fig. 2.6), except that gates and auxiliary strands are scaled by a factor of s. Further, backward auxiliary species are added at $s \cdot 2$x concentration. Strand displacement-level rate constants are from section S7. The concentrations of the signal species are calculated from PX, PY, and PB concentrations as in the main text.

4.2 In vitro mixing test under different experimental regimes

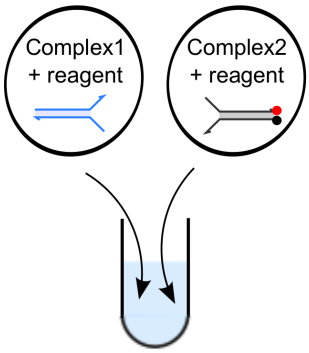
While co-delivery of nucleic acids with lipofectamine is experimentally easy, gates can possibly interact before entering cells. We performed the *in vitro* mixing experiments to investigate this, and we found that several parameters need to be carefully optimized to minimize the "leak" reactions between the reporter and input before they enter the cell.

First, the use of phenol red media could dramatically interfere with the reading of the fluorescence. In contrast to using clear media, an unexpected signal drop of the activated reporter (positive control) was observed when experiments were performed in media with phenol red (Fig. 4.3). This interference was likely caused by the fluorescence bleed-through from red (phenol-red media) to far-red (our Cy5 labeled reporter).

Second, this leaky reaction was very sensitive to the stoichiometry ratio of the transfection reagent to the nucleic acid cargo. We observed that increasing this stoichiometry ratio could efficiently reduce the leak (Fig. 4.4). Note that the signal of activated reporter (positive control) decreased when more transfection reagents were used. This is likely because the lipofectamin/nucleic acid transfection complex facilitated nucleic acid condensation, and the fluorophores could quench each other when they were in close proximity.

Lastly, we varied the stoichiometry ratios of L2K to the input to optimize the experimental regime. Interestingly, the leakage appeared to be approximately the same regardless of 3 or 6 μL of L2K was used for 150 nM of the input (Fig. 4.5). While increasing the amount of L2K for packaging the input did not reduce the leakage, the transfection reagent was found to be toxic to cells. To strike a balance between leakage and cell viability, we used 3 μL of L2K for 150 nM of the input and 2 μL of L2K for 50 nM of the reporter for all our co-transfection experiments (except where explicitly mentioned).

a



	Complex1	Complex2
Positive control	active reporter	scrambled input
Negative control	reporter	scrambled input
Test case	reporter	input
Positive control (Pre-mixed)	reporter	input

b

i, Opti-MEM with pheno red

ii, Opti-MEM without pheno red

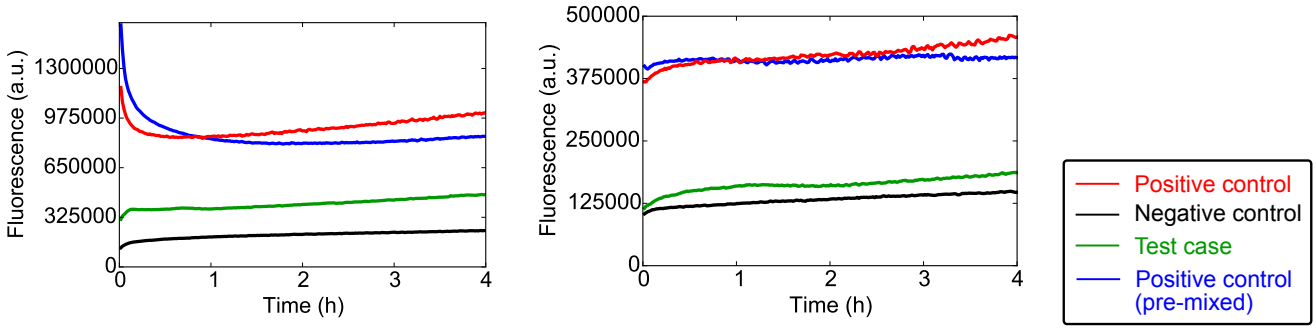


Figure 4.3: *In vitro* mixing test between medium with and without phenol red. Table **a** shows the complexes used for the corresponding experiments. **b,** *In vitro* kinetics. All complexes were at 50 nM. The blue trace corresponds to a positive control where the reporter and input were mixed with each other at room temperature (RT) for 20 minutes. Subsequently, 4 μL of Lipofectamine2000 (L2K) was added to this mixture before tracking the fluorescence. Other traces correspond to experiments where each complex was incubated with 2 μL of L2K independently at RT for 20 minutes. Immediately after mixing complex1 and complex2, the fluorescence signal was tracked using spectrometer in a cell-free setting at 37°C. Experiments were performed in (i) Opti-MEM with phenol red or (ii) Opti-MEM without phenol red.

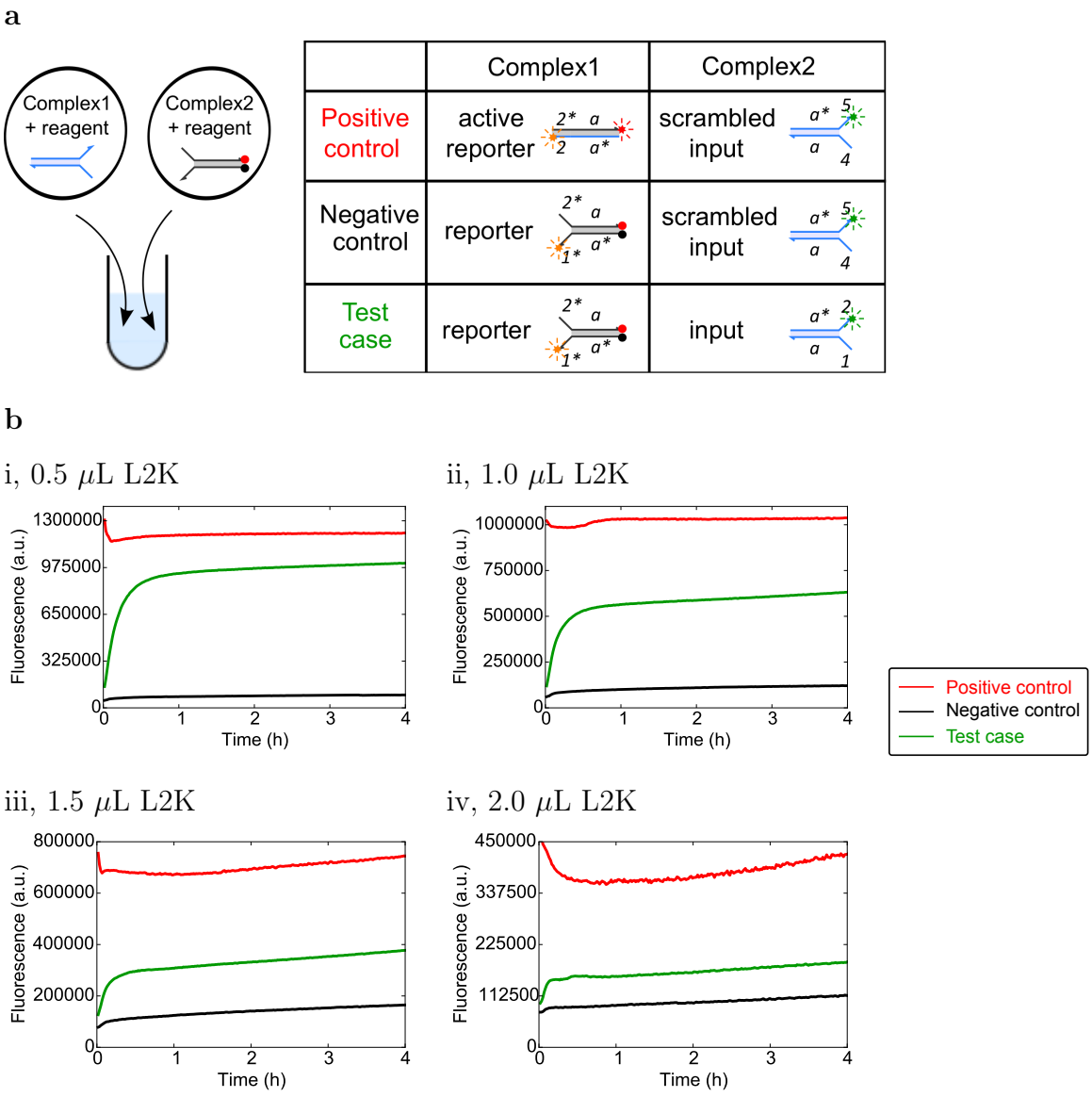


Figure 4.4: *In vitro* mixing test using different stoichiometry ratios of the transfection reagent to the nucleic acid cargo. Table **a** shows the complexes used for the corresponding experiments. **b,** *In vitro* kinetics. Complex1 was labeled with Cy5 (red). Constitutively active fluorophores TYE563 (orange) and ATTO488 (green) were added to complex1 and complex2, respectively, but these two fluorophores were only used to adjust the variation of in-cell experiments. All complexes were at 50 nM. Each complex was independently incubated with L2K at RT for 20 minutes (the amount of L2K is indicated in the figure). Immediately after mixing these two complexes, the fluorescence signal was tracked using spectrometer in a cell-free setting. The experiments were performed in Opti-MEM without phenol red at 37°C.

a

	Complex1		Complex2	
Positive control	active reporter	2uL L2K	scrambled input	3uL L2K
Negative control	reporter	2uL L2K	scrambled input	3uL L2K
Test case1	reporter	2uL L2K	input	3uL L2K
Test case2	reporter	2uL L2K	input	6uL L2K

b

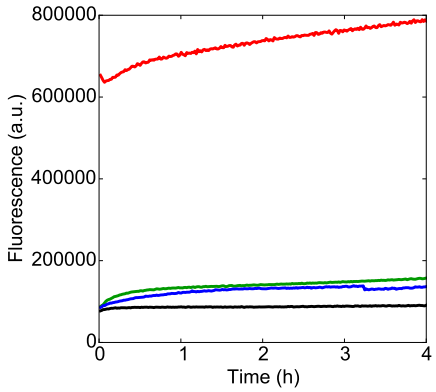


Figure 4.5: *In vitro* mixing test with different stoichiometry ratios of L2K to the input. Table **a** shows the complexes and the amount of L2K used for the corresponding experiments. **b,** *In vitro* kinetics. Complex1 was labeled with Cy5 (red). Constitutively active fluorophores TYE563 (orange) and ATTO488 (green) were added to complex1 and complex2, respectively, but these two fluorophores were only used to adjust the variation of in-cell experiments. All complex1 was at 50 nM and all complex2 was at 150 nM. Each complex was incubated with L2K separately, and only subsequently mixed. The experiments were performed in Opti-MEM without phenol red at 37°C, and the fluorescence signal was tracked using spectrometer.

4.3 Detecting cellular mRNA with strand displacement probes.

Strand displacement circuits have a natural pathway for diagnosing diseases *in vivo*: recognizing their RNA profiles. However, detecting RNA in live cells is a difficult challenge because of their low copy number, and also the requirement of nanodevices to access the cytoplasm, where most mature mRNA are located.

Here, we took first steps towards detecting endogenous mRNA in live cells with strand displacement (SD) probes. We began by testing the mRNA hybridization *in situ*. Then we used molecular beacons—a class of well studied probes for mRNA imaging inside cells—to detect endogenous mRNA in living cells. Lastly, we engineered a variety of SD probes to investigate whether structures affect their ability of detecting mRNA in living cells.

4.3.1 In situ detection of repeated sequences in individual mRNA molecules.

To achieve single-molecule sensitivity, we used mammalian HT-1080 cells that were previously engineered to express RNA transcripts with 96-tandem repeats in its 3 untranslated region (this cell line is referred as HT-1080-96X)[66]. These tandemly repeated sequences allow up to 96 fluorescence probes to bind to the same transcript, which can greatly improve the signal to background ratio. Indeed, using fluorescence *in situ* hybridization (FISH), single transcripts expressed in HT-1080-96X became detectable as single bright fluorescent spots (Fig. 4.6). In contrast, the control cell line showed only background fluorescence.

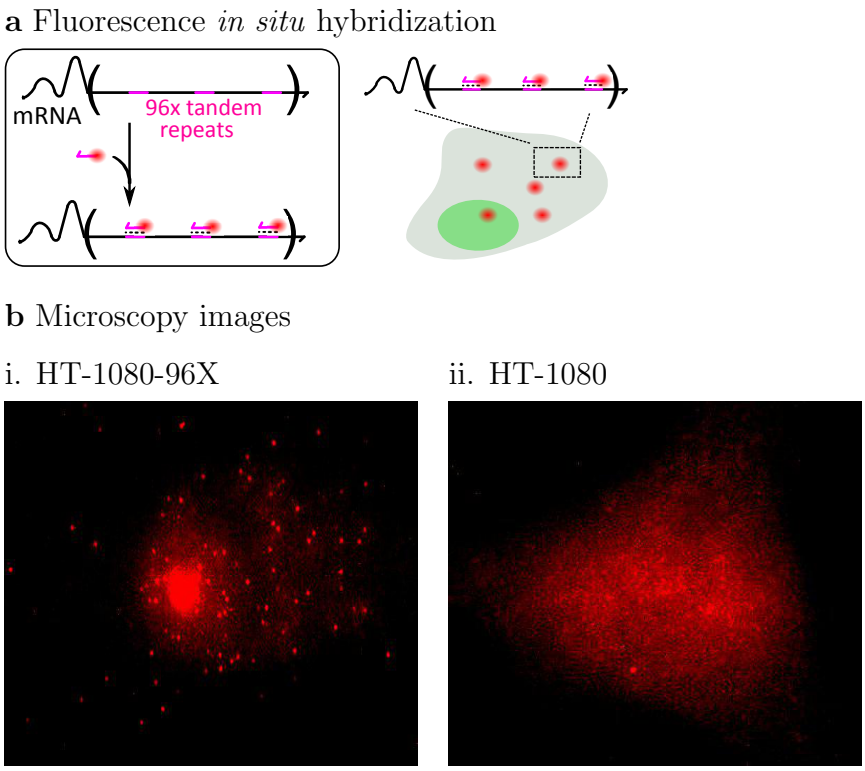


Figure 4.6: *In situ* detecting of repeated sequences in individual mRNA molecules. a, Mechanism of fluorescence *in situ* hybridization. Multiple fluorophore labeled strands can hybridize to the tandem repeats in the 3'UTR of target mRNA in HT-1080-96X. **b,** Representative fluorescent microscopy images. Identifiable spots were observed in HT-1080-96X (i) but not in HT-1080 (ii, control cell line).

4.3.2 Detection of mRNA using molecular beacons in live cells.

We then used a live-cell imaging probe, called ratiometric molecular beacon (RBMB) [66], to test whether this modified gene expressed in HT-1080-96X are detectable in live cells. These probes contained initially quenched fluorphores that can be restored upon hybridization to a specific target sequence (Fig. 4.7a). Moreover, the RBMBs contained addition of longer double-stranded domain, which facilitates their export from the nucleus to the cytosol. After electroporation, live cell images showed that individual spots could be identified in HT-1080-96X cells, indicating multiple RBMBs were bound to the 3'UTR regions (Fig. 4.7b).

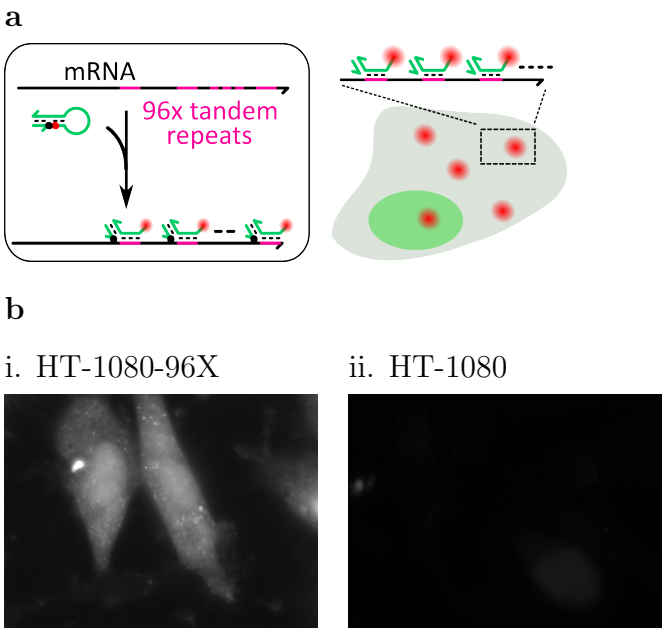


Figure 4.7: Endogenous mRNA detection using ratiometric bimolecular beacons (RBMBs). a, Binding to a target mRNA separates the reporter dye (red dot) from the quencher (black dot), resulting in high fluorescence. Multiple RBMBs can bind to the tandem repeat targets in the 3'UTR of an mRNA, enabling visualization of a single transcript in living cells. **b,** Representative microscopy images (Cy5 channel). Images were taken two and a half hours following electroporation. i. Identifiable fluorescent spots were observed in HT-1080-96X cells. ii. Cells without any target sites (HT-1080) only showed minimal background.

4.3.3 Detecting cellular mRNA with strand displacement probes.

To study whether the structures of probes affect the performance of the detection of mRNA, we engineered strand displacement (SD) based probes with different lengths of toeholds and double-stranded (ds) regions (Fig. 4.8). The SD probes can hybridize to the target sequence via toeholds, and then displace the top quencher-labeled strand, resulting in high fluorescence.

We began by characterizing these SD probes in cell-free settings using synthesized RNA targets with sequences equal to the tandem repeats. Interestingly, we observed that our probes could be sped up by increasing the toehold length and/or shortening the length of the double-stranded regions. This result is different from previously reported SD probes that reach the maximum speed when the toehold length is over 8-mer. Further characterization of RNA targets with secondary structures may improve the understanding of this reaction. Also note that the background fluorescence of RBMB was higher than that of SD reporters, which is likely because the hairpin structure of RBMB was metastable and the fluorophore was partially quenched.

We then electroporated SD probes into HT-1080-96X cells, and took microscopy images to see whether the probes can detect mRNA targets. We found that the sub-cellular localization strongly depended on the probe structure: only the probe with 18-bp was able to access the cytosol (Fig. 4.8c). However, no crispy spots were identified after the delivery of this 18-bp SD probe. The failure of this SD probe might be caused by its slower reaction speed compared to RBMB. The problem can be mitigated by elongating the toehold length, which base been shown to effectly increase the reaction speed.

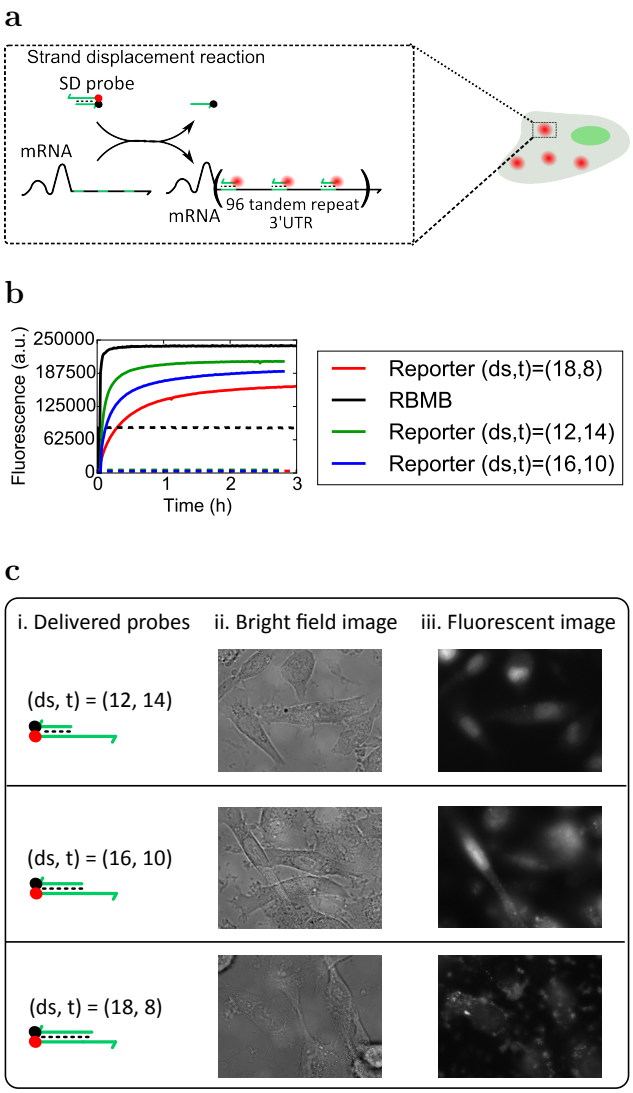


Figure 4.8: Detection of endogenous mRNA using strand displacement (SD) probes. **a,** Designed pathway of SD probes. mRNA target sites can interact with SD probes, displacing the quencher (black dot) labeled strand and resulting in high fluorescence. The binding of SD probes to multiple target sites can enable the visualization of a single mRNA in live cells. **b,** *In vitro* time-course data. Here the sequences of the RNA target is equal to each tandem repeat (i.e. 50 bp). "Reporter only" was shown as dashed traces; "reporter+mRNA target" was shown as solid traces. The toehold length t and length of double-stranded region ds are indicated in the legend. Reporter was at 90 nM and the RNA target was at 30 nM. Reporter was 2'Ome RNA. The experiments were performed in 1X PBS at 37°C. **c,** Representative microscopy images. HT-1080-96X cells were delivered with different SD probes. Panels (i) show the delivered probes. Following two hours of probe delivery, live cell images were taken. Bright-field images and fluorescent images are shown in panels (ii) and (iii), respectively.

4.3.4 Methods: electroporation.

The following electroporation protocol is from [129]. I added additional notes to steps that require particular attention.

1. Aspirate growth media from the engineered cells (60-80% confluent), and wash once with 1X DPBS (Ca^{2+} and Mg^{2+} free).

2. Add 1 mL trypsin and incubate at 37°C for 2-3 minutes (min). Stop the trypsinization by adding 1 mL DMEM/Opti-MEM supplemented with 10% fetal bovine serum(FBS) without antibiotics and phenol red.

3. Spin down the cells at 200∗g for 5 min.

4. Aspirate the supernatant and resuspend the cells using 1 mL DPBS.

5. Count the cells using hemocytometer or flow cytometer.

6. Transfer 600,000 cells with DPBS into a 1.5 mL tube and spin down at 200∗g for 5 min.

7. Aspirate the supernatant, and add 22 μL resuspension buffer.

8. Add 2 μL of the probe at 12 μM, and mix it by pipetting up and down several times. Be careful not to make air bubbles.

9. Microporate 10 μL of the probe-cell mixture. Note that it is very important to avoid air bubbles during electroporation. Otherwise, the probe delivery will not work.

10. Expel the 10 μL of the electroplated mixture into 1 mL Opti-MEM/DMEM with FBS but without antibiotics and phenol-red. Mix the sample by inverting it several times.

11. Spin down the cells at 200∗g for 5 min.

12. Remove the supernatant.

13. Resuspend the cell pellet using 1 mL Opti-MEM/DMEM with FBS but without antibiotics and phenol-red.

14. Wash the cells twice by repeating steps 11-13.

15. Resuspend the cells with 400 μL media with FBS but without antibiotics and phenol red.

16. Plate the cells on a 24-well ECM-coated plate.

17. Incubate at 37° with 5% CO_2 for at least 1 hour before imaging.

Note: Although step 10 only uses 10 μL of the probe-cell mixture, it is important to prepare excess amounts in steps 6-8 because it can help avoid generating air bubbles while using microporation pipette.

BIBLIOGRAPHY

[1] Donald M Crothers, Victor A Bloomfield, and Ignacio Tinoco. *Nucleic acids: structures, properties, and functions*. University science books, 2000.

[2] John SantaLucia Jr and Donald Hicks. The thermodynamics of dna structural motifs. *Annu. Rev. Biophys. Biomol. Struct.*, 33:415–440, 2004.

[3] Robert Carlson. The changing economics of dna synthesis. *Nature biotechnology*, 27(12):1091–1094, 2009.

[4] Wendy U Dittmer, Andreas Reuter, and Friedrich C Simmel. A dna-based machine that can cyclically bind and release thrombin. *Angewandte Chemie International Edition*, 43(27):3550–3553, 2004.

[5] Bernard Yurke, Allen P Mills, and Siu Lai Cheng. Dna implementation of addition in which the input strands are separate from the operator strands. *Biosystems*, 52(1):165–174, 1999.

[6] Yaakov Benenson, Binyamin Gil, Uri Ben-Dor, Rivka Adar, and Ehud Shapiro. An autonomous molecular computer for logical control of gene expression. *Nature*, 429(6990):423–429, 2004.

[7] SeungHyeon Ko, Haipeng Liu, Yi Chen, and Chengde Mao. Dna nanotubes as combinatorial vehicles for cellular delivery. *Biomacromolecules*, 9(11):3039–3043, 2008.

[8] Nadrian C Seeman. Nanomaterials based on dna. *Annual review of biochemistry*, 79:65, 2010.

[9] Jeanette Nangreave, Dongran Han, Yan Liu, and Hao Yan. Dna origami: a history and current perspective. *Current opinion in chemical biology*, 14(5):608–615, 2010.

[10] Akinori Kuzuya and Makoto Komiyama. Dna origami: fold, stick, and beyond. *Nanoscale*, 2(3):309–321, 2010.

[11] Andre V Pinheiro, Dongran Han, William M Shih, and Hao Yan. Challenges and opportunities for structural dna nanotechnology. *Nature nanotechnology*, 6(12):763–772, 2011.

[12] Thomas Tørring, Niels V Voigt, Jeanette Nangreave, Hao Yan, and Kurt V Gothelf. Dna origami: a quantum leap for self-assembly of complex structures. *Chemical Society Reviews*, 40(12):5636–5646, 2011.

[13] David Yu Zhang and Georg Seelig. Dynamic dna nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3(2):103–113, 2011.

[14] Leonard M Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.

[15] Bernard Yurke, Andrew J Turberfield, Allen P Mills, Friedrich C Simmel, and Jennifer L Neumann. A dna-fuelled molecular machine made of dna. *Nature*, 406(6796):605–608, 2000.

[16] Qingge Li, Guoyan Luan, Qiuping Guo, and Jixuan Liang. A new class of homogeneous nucleic acid probes based on specific displacement hybridization. *Nucleic Acids Research*, 30(2):e5–e5, 2002.

[17] Bernard Yurke and Allen P Mills Jr. Using dna to power nanostructures. *Genetic Programming and Evolvable Machines*, 4(2):111–122, 2003.

[18] David Yu Zhang and Erik Winfree. Control of dna strand displacement kinetics using toehold exchange. *Journal of the American Chemical Society*, 131(47):17303–17314, 2009.

[19] Niranjan Srinivas, Thomas E Ouldridge, Petr Šulc, Joseph M Schaeffer, Bernard Yurke, Ard A Louis, Jonathan PK Doye, and Erik Winfree. On the biophysics and kinetics of toehold-mediated dna strand displacement. *Nucleic acids research*, 41(22):10641–10658, 2013.

[20] Georg Seelig, David Soloveichik, David Yu Zhang, and Erik Winfree. Enzyme-free nucleic acid logic circuits. *science*, 314(5805):1585–1588, 2006.

[21] Robert M Dirks and Niles A Pierce. Triggered amplification by hybridization chain reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43):15275–15278, 2004.

[22] William B Sherman and Nadrian C Seeman. A precisely controlled dna biped walking device. *Nano Letters*, 4(7):1203–1207, 2004.

[23] Jong-Shik Shin and Niles A Pierce. A synthetic dna walker for molecular transport. *Journal of the American Chemical Society*, 126(35):10834–10835, 2004.

[24] Hongzhou Gu, Jie Chao, Shou-Jun Xiao, and Nadrian C Seeman. A proximity-based programmable dna nanoscale assembly line. *Nature*, 465(7295):202–205, 2010.

[25] Peng Yin, Harry MT Choi, Colby R Calvert, and Niles A Pierce. Programming biomolecular self-assembly pathways. *Nature*, 451(7176):318–322, 2008.

[26] Tosan Omabegho, Ruojie Sha, and Nadrian C Seeman. A bipedal dna brownian motor with coordinated legs. *Science*, 324(5923):67–71, 2009.

[27] Paul WK Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[28] Kyle Lund, Anthony J Manzo, Nadine Dabby, Nicole Michelotti, Alexander Johnson-Buck, Jeanette Nangreave, Steven Taylor, Renjun Pei, Milan N Stojanovic, Nils G Walter, et al. Molecular robots guided by prescriptive landscapes. *Nature*, 465(7295):206–210, 2010.

[29] Shelley FJ Wickham, Masayuki Endo, Yousuke Katsuda, Kumi Hidaka, Jonathan Bath, Hiroshi Sugiyama, and Andrew J Turberfield. Direct observation of stepwise movement of a synthetic molecular transporter. *Nature nanotechnology*, 6(3):166–169, 2011.

[30] Shelley FJ Wickham, Jonathan Bath, Yousuke Katsuda, Masayuki Endo, Kumi Hidaka, Hiroshi Sugiyama, and Andrew J Turberfield. A dna-based molecular motor that can navigate a network of tracks. *Nature nanotechnology*, 7(3):169–173, 2012.

[31] Yaakov Benenson, Tamar Paz-Elizur, Rivka Adar, Ehud Keinan, Zvi Livneh, and Ehud Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 414(6862):430–434, 2001.

[32] Milan N Stojanovic, Tiffany Elizabeth Mitchell, and Darko Stefanovic. Deoxyribozyme-based logic gates. *Journal of the American Chemical Society*, 124(14):3555–3561, 2002.

[33] Renjun Pei, Elizabeth Matamoros, Manhong Liu, Darko Stefanovic, and Milan N Stojanovic. Training a molecular automaton to play a game. *Nature nanotechnology*, 5(11):773–777, 2010.

[34] David Soloveichik, Georg Seelig, and Erik Winfree. Dna as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, 2010.

[35] Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from dna. *Nature nanotechnology*, 8(10):755–762, 2013.

[36] Kevin Oishi and Eric Klavins. Biomolecular implementation of linear i/o systems. *Systems Biology, IET*, 5(4):252–260, 2011.

[37] Andrew J Turberfield, JC Mitchell, Bernard Yurke, AP Mills, MI Blakey, and Friedrich C Simmel. Dna fuel for free-running nanomachines. *Physical review letters*, 90(11):118102–118102, 2003.

[38] Justin S Bois, Suvir Venkataraman, Harry MT Choi, Andrew J Spakowitz, Zhen-Gang Wang, and Niles A Pierce. Topological constraints in nucleic acid hybridization kinetics. *Nucleic acids research*, 33(13):4090–4095, 2005.

[39] Simon J Green, Daniel Lubrich, and Andrew J Turberfield. Dna hairpins: fuel for autonomous dna devices. *Biophysical journal*, 91(8):2966–2975, 2006.

[40] Georg Seelig, Bernard Yurke, and Erik Winfree. Catalyzed relaxation of a metastable dna fuel. *Journal of the American Chemical Society*, 128(37):12211–12220, 2006.

[41] David Yu Zhang, Andrew J Turberfield, Bernard Yurke, and Erik Winfree. Engineering entropy-driven reactions and networks catalyzed by dna. *Science*, 318(5853):1121–1125, 2007.

[42] Lulu Qian and Erik Winfree. Scaling up digital circuit computation with dna strand displacement cascades. *Science*, 332(6034):1196–1201, 2011.

[43] Xi Chen, Neima Briggs, Jeremy R McLain, and Andrew D Ellington. Stacking nonenzymatic circuits for high signal gain. *Proceedings of the National Academy of Sciences*, 110(14):5386–5391, 2013.

[44] Ebbe S Andersen, Mingdong Dong, Morten M Nielsen, Kasper Jahn, Ramesh Subramani, Wael Mamdouh, Monika M Golas, Bjoern Sander, Holger Stark, Cristiano LP Oliveira, et al. Self-assembly of a nanoscale dna box with a controllable lid. *Nature*, 459(7243):73–76, 2009.

[45] Harry MT Choi, Joann Y Chang, Le A Trinh, Jennifer E Padilla, Scott E Fraser, and Niles A Pierce. Programmable in situ amplification for multiplexed imaging of mrna expression. *Nature biotechnology*, 28(11):1208–1212, 2010.

[46] Harry MT Choi, Victor A Beck, and Niles A Pierce. Next-generation in situ hybridization chain reaction: higher gain, lower cost, greater durability. *ACS nano*, 8(5):4284–4294, 2014.

[47] Marshall J Levesque, Paul Ginart, Yichen Wei, and Arjun Raj. Visualizing snvs to quantify allele-specific expression in single cells. *Nature methods*, 10(9):865–867, 2013.

[48] Arjun Raj, Patrick Van Den Bogaard, Scott A Rifkin, Alexander Van Oudenaarden, and Sanjay Tyagi. Imaging individual mrna molecules using multiple singly labeled probes. *Nature methods*, 5(10):877–879, 2008.

[49] Dzifa Y Duose, Ryan M Schweller, Jan Zimak, Arthur R Rogers, Walter N Hittelman, and Michael R Diehl. Configuring robust dna strand displacement reactions for in situ molecular analyses. *Nucleic acids research*, 40(7):3289–3298, 2012.

[50] Ralf Jungmann, Christian Steinhauer, Max Scheible, Anton Kuzyk, Philip Tinnefeld, and Friedrich C Simmel. Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on dna origami. *Nano letters*, 10(11):4756–4761, 2010.

[51] Ralf Jungmann, Maier S Avendaño, Johannes B Woehrstein, Mingjie Dai, William M Shih, and Peng Yin. Multiplexed 3d cellular super-resolution imaging with dna-paint and exchange-paint. *Nature methods*, 11(3):313–318, 2014.

[52] Anthony D Keefe, Supriya Pai, and Andrew Ellington. Aptamers as therapeutics. *Nature Reviews Drug Discovery*, 9(7):537–550, 2010.

[53] Shawn M Douglas, Ido Bachelet, and George M Church. A logic-gated nanorobot for targeted transport of molecular payloads. *Science*, 335(6070):831–834, 2012.

[54] Craig Tuerk and Larry Gold. Systematic evolution of ligands by exponential enrichment: Rna ligands to bacteriophage t4 dna polymerase. *Science*, 249(4968):505–510, 1990.

[55] Andrew D Ellington and Jack W Szostak. In vitro selection of rna molecules that bind specific ligands. *nature*, 346(6287):818–822, 1990.

[56] Yaniv Amir, Eldad Ben-Ishay, Daniel Levner, Shmulik Ittah, Almogit Abu-Horowitz, and Ido Bachelet. Universal computing by dna origami robots in a living animal. *Nature nanotechnology*, 9(5):353–357, 2014.

[57] Mingxu You, Lu Peng, Na Shao, Liqin Zhang, Liping Qiu, Cheng Cui, and Weihong Tan. Dna nano-claw: logic-based autonomous cancer targeting and therapy. *Journal of the American Chemical Society*, 136(4):1256–1259, 2014.

[58] Mingxu You, Guizhi Zhu, Tao Chen, Michael J Donovan, and Weihong Tan. Programmable and multiparameter dna-based logic platform for cancer recognition and targeted therapy. *Journal of the American Chemical Society*, 137(2):667–674, 2014.

[59] Maria Rudchenko, Steven Taylor, Payal Pallavi, Alesia Dechkovskaia, Safana Khan, Vincent P Butler Jr, Sergei Rudchenko, and Milan N Stojanovic. Autonomous molecular cascades for evaluation of cell surfaces. *Nature nanotechnology*, 8(8):580–586, 2013.

[60] Hao Pei, Le Liang, Guangbao Yao, Jiang Li, Qing Huang, and Chunhai Fan. Reconfigurable three-dimensional dna nanostructures for the construction of intracellular logic sensors. *Angewandte Chemie*, 124(36):9154–9158, 2012.

[61] Souvik Modi, MG Swetha, Debanjan Goswami, Gagan D Gupta, Satyajit Mayor, and Yamuna Krishnan. A dna nanomachine that maps spatial and temporal ph changes inside living cells. *Nature Nanotechnology*, 4(5):325–330, 2009.

[62] Souvik Modi, Clément Nizak, Sunaina Surana, Saheli Halder, and Yamuna Krishnan. Two dna nanomachines map ph changes along intersecting endocytic pathways inside the same cell. *Nature nanotechnology*, 2013.

[63] Sanjay Tyagi, Fred Russell Kramer, et al. Molecular beacons: probes that fluoresce upon hybridization. *Nature biotechnology*, 14(3):303–308, 1996.

[64] Antony K Chen, Olga Davydenko, Mark A Behlke, and Andrew Tsourkas. Ratiometric bimolecular beacons for the sensitive detection of rna in single living cells. *Nucleic acids research*, page gkq436, 2010.

[65] Musa M Mhlanga, Diana Y Vargas, Cindy W Fung, Fred Russell Kramer, and Sanjay Tyagi. trna-linked molecular beacons for imaging mrnas in the cytoplasm of living cells. *Nucleic acids research*, 33(6):1902–1912, 2005.

[66] Xuemei Zhang, Yang Song, Akash Y Shah, Virzhiniya Lekova, Arjun Raj, Ling Huang, Mark A Behlke, and Andrew Tsourkas. Quantitative assessment of ratiometric bimolecular beacons as a tool for imaging single engineered rna transcripts and measuring gene expression in living cells. *Nucleic acids research*, 41(15):e152–e152, 2013.

[67] Diana P Bratu, Byeong-Jik Cha, Musa M Mhlanga, Fred Russell Kramer, and Sanjay Tyagi. Visualizing the distribution and transport of mrnas in living cells. *Proceedings of the National Academy of Sciences*, 100(23):13308–13313, 2003.

[68] Philip J Santangelo, Aaron W Lifland, Paul Curt, Yukio Sasaki, Gary J Bassell, Michael E Lindquist, and James E Crowe. Single molecule–sensitive probes for imaging rna in live cells. *Nature methods*, 6(5):347–349, 2009.

[69] Nathaniel L Rosi, David A Giljohann, C Shad Thaxton, Abigail KR Lytton-Jean, Min Su Han, and Chad A Mirkin. Oligonucleotide-modified gold nanoparticles for intracellular gene regulation. *Science*, 312(5776):1027–1030, 2006.

[70] Ali H Alhasan, Pinal C Patel, Chung Hang J Choi, and Chad A Mirkin. Exosome encased spherical nucleic acid gold nanoparticle conjugates as potent microrna regulation agents. *Small*, 10(1):186–192, 2014.

[71] Andrew E Prigodich, Dwight S Seferos, Matthew D Massich, David A Giljohann, Brandon C Lane, and Chad A Mirkin. Nano-flares for mrna regulation and detection. *Acs Nano*, 3(8):2147–2152, 2009.

[72] Tiffany L Halo, Kaylin M McMahon, Nicholas L Angeloni, Yilin Xu, Wei Wang, Alyssa B Chinen, Dmitry Malin, Elena Strekalova, Vincent L Cryns, Chonghui Cheng, et al. Nanoflares for the detection, isolation, and culture of live tumor cells from human blood. *Proceedings of the National Academy of Sciences*, 111(48):17104–17109, 2014.

[73] Kirill A Afonin, Mathias Viard, Angelica N Martins, Stephen J Lockett, Anna E Maciag, Eric O Freed, Eliahu Heldman, Luc Jaeger, Robert Blumenthal, and Bruce A Shapiro. Activation of different split functionalities on re-association of rna-dna hybrids. *Nature nanotechnology*, 2013.

[74] Sherry Xi Chen, David Yu Zhang, and Georg Seelig. Conditionally fluorescent molecular probes for detecting single base changes in double-stranded dna. *Nature chemistry*, 5(9):782–789, 2013.

[75] Zhen Xie, Siyuan John Liu, Leonidas Bleris, and Yaakov Benenson. Logic integration of mrna signals by an rnai-based molecular computer. *Nucleic acids research*, 38(8):2692–2701, 2010.

[76] Lisa M Hochrein, Maayan Schwarzkopf, Mona Shahgholi, Peng Yin, and Niles A Pierce. Conditional dicer substrate formation via shape and sequence transduction with small conditional rnas. *Journal of the American Chemical Society*, 135(46):17322–17330, 2013.

[77] Deepak Kumar, Sang Hoon Kim, and Yohei Yokobayashi. Combinatorially inducible rna interference triggered by chemically modified oligonucleotides. *Journal of the American Chemical Society*, 133(8):2783–2788, 2011.

[78] Maya Kahan-Hanum, Yehonatan Douek, Rivka Adar, and Ehud Shapiro. A library of programmable dnazymes that operate in a cellular environment. *Scientific reports*, 3, 2013.

[79] James Hemphill and Alexander Deiters. Dna computation in mammalian cells: microrna logic operations. *Journal of the American Chemical Society*, 135(28):10512–10518, 2013.

[80] Luca Cardelli. Two-domain dna strand displacement. *Math. Struct. Comput. Sci.*, 23:247–271, 2013.

[81] K Eric Drexler. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proceedings of the National Academy of Sciences*, 78(9):5275–5278, 1981.

[82] Otilia M Koo, Israel Rubinstein, and Hayat Onyuksel. Role of nanotechnology in targeted drug delivery and imaging: a concise review. *Nanomedicine: Nanotechnology, Biology and Medicine*, 1(3):193–212, 2005.

[83] Henry Hess. Engineering applications of biomolecular motors. *Annual review of biomedical engineering*, 13:429–450, 2011.

[84] Matthew Levy and Andrew D Ellington. Exponential growth by cross-catalytic cleavage of deoxyribozymogens. *Proceedings of the National Academy of Sciences*, 100(11):6416–6421, 2003.

[85] Andrew Phillips and Luca Cardelli. A programming language for composable dna circuits. *Journal of the Royal Society Interface*, 6(Suppl 4):S419–S436, 2009.

[86] Lulu Qian, David Soloveichik, and Erik Winfree. Efficient turing-universal computation with dna polymers. In *DNA computing and molecular programming*, pages 123–140. Springer, 2011.

[87] Lulu Qian, Erik Winfree, and Jehoshua Bruck. Neural network computation with dna strand displacement cascades. *Nature*, 475(7356):368–372, 2011.

[88] Milan N Stojanovic and Darko Stefanovic. A deoxyribozyme-based molecular automaton. *Nature biotechnology*, 21(9):1069–1074, 2003.

[89] Kevin Montagne, Raphael Plasson, Yasuyuki Sakai, Teruo Fujii, and Yannick Rondelez. Programming an in vitro dna oscillator using a molecular networking strategy. *Molecular systems biology*, 7(1), 2011.

[90] Itamar Willner, Bella Shlyahovsky, Maya Zayats, and Bilha Willner. Dnazymes for sensing, nanobiotechnology and logic gate applications. *Chemical Society Reviews*, 37(6):1153–1165, 2008.

[91] Tom Ran, Shai Kaplan, and Ehud Shapiro. Molecular implementation of simple logic programs. *Nature Nanotechnology*, 4(10):642–648, 2009.

[92] Richard A Muscat, Jonathan Bath, and Andrew J Turberfield. A programmable molecular robot. *Nano letters*, 11(3):982–987, 2011.

[93] Erik Winfree, Furong Liu, Lisa A Wenzler, and Nadrian C Seeman. Design and self-assembly of two-dimensional dna crystals. *Nature*, 394(6693):539–544, 1998.

[94] Irving R Epstein, John A Pojman, and Gregoire Nicolis. An introduction to nonlinear chemical dynamics: oscillations, waves, patterns, and chaos. *Physics Today*, 52:68, 1999.

[95] Marcelo O Magnasco. Chemical kinetics is turing universal. *Physical Review Letters*, 78(6):1190–1193, 1997.

[96] Phillip Senum and Marc Riedel. Rate-independent constructs for chemical computation. *PloS one*, 6(6):e21414, 2011.

[97] Petri net theory and the modeling of systems. 1981.

[98] David Yu Zhang and Erik Winfree. Robustness and modularity properties of a non-covalent dna catalytic reaction. *Nucleic acids research*, 38(12):4182–4197, 2010.

[99] Chenxiang Lin, Sherri Rinker, Xing Wang, Yan Liu, Nadrian C Seeman, and Hao Yan. In vivo cloning of artificial dna nanostructures. *Proceedings of the National Academy of Sciences*, 105(46):17626–17631, 2008.

[100] Cosimo Ducani, Corinna Kaul, Martin Moche, William M Shih, and Björn Högberg. Enzymatic production of'monoclonal stoichiometric'single-stranded dna oligonucleotides. *Nature methods*, 2013.

[101] Dana Angluin, James Aspnes, and David Eisenstat. A simple population protocol for fast robust approximate majority. *Distributed Computing*, 21(2):87–102, 2008.

[102] Matthew R Lakin, Simon Youssef, Luca Cardelli, and Andrew Phillips. Abstractions for dna circuit design. *Journal of The Royal Society Interface*, 9(68):470–486, 2012.

[103] Luca Cardelli and Attila Csikász-Nagy. The cell cycle switch computes approximate majority. *Scientific reports*, 2, 2012.

[104] David Yu Zhang and Georg Seelig. Dna-based fixed gain amplifiers and linear classifier circuits. In *DNA Computing and Molecular Programming*, pages 176–186. Springer, 2011.

[105] P ROBERT Christian and George Casella. Monte carlo statistical methods, 1999.

[106] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature methods*, 6(5):343–345, 2009.

[107] Johann Elbaz, Oleg Lioubashevski, Fuan Wang, Françoise Remacle, Raphael D Levine, and Itamar Willner. Dna computing circuits using libraries of dnazyme subunits. *Nature nanotechnology*, 5(6):417–422, 2010.

[108] Suvir Venkataraman, Robert M Dirks, Paul WK Rothemund, Erik Winfree, and Niles A Pierce. An autonomous polymerization motor powered by dna hybridization. *Nature Nanotechnology*, 2(8):490–494, 2007.

[109] SJ Green, Jonathan Bath, and AJ Turberfield. Coordinated chemomechanical cycles: a mechanism for autonomous molecular motion. *Physical review letters*, 101(23):238101, 2008.

[110] David Yu Zhang, Rizal F Hariadi, Harry MT Choi, and Erik Winfree. Integrating dna strand-displacement circuitry with dna tile self-assembly. *Nature communications*, 4, 2013.

[111] Qiao Jiang, Chen Song, Jeanette Nangreave, Xiaowei Liu, Lin Lin, Dengli Qiu, Zhen-Gang Wang, Guozhang Zou, Xingjie Liang, Hao Yan, et al. Dna origami as a carrier for circumvention of drug resistance. *Journal of the American Chemical Society*, 134(32):13396–13403, 2012.

[112] Verena J Schller, Simon Heidegger, Nadja Sandholzer, Philipp C Nickels, Nina A Suhartha, Stefan Endres, Carole Bourquin, and Tim Liedl. Cellular immunostimulation by cpg-sequence-coated dna origami structures. *ACS nano*, 5(12):9696–9702, 2011.

[113] Camille J Delebecque, Ariel B Lindner, Pamela A Silver, and Faisal A Aldaye. Organization of intracellular reactions with rationally designed rna assemblies. *Science*, 333(6041):470–474, 2011.

[114] William K Holloman, Roger Wiegand, Catherine Hoessli, and Charles M Radding. Uptake of homologous single-stranded fragments by superhelical dna: a possible mechanism for initiation of genetic recombination. *Proceedings of the National Academy of Sciences*, 72(6):2394–2398, 1975.

[115] Betty J Thompson, Merrill N Camien, and Robert C Warner. Kinetics of branch migration in double-stranded dna. *Proceedings of the National Academy of Sciences*, 73(7):2299–2303, 1976.

[116] Philip J Santangelo. Molecular beacons and related probes for intracellular rna imaging. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*, 2(1):11–19, 2010.

[117] Gang Bao, Won Jong Rhee, and Andrew Tsourkas. Fluorescent probes for live-cell rna detection. *Annual review of biomedical engineering*, 11:25, 2009.

[118] Jesper B Bramsen, Maria B Laursen, Anne F Nielsen, Thomas B Hansen, Claus Bus, Niels Langkjær, B Ravindra Babu, Torben Højland, Mikhail Abramov, Arthur Van Aerschot, et al. A large-scale chemical modification screen identifies design rules to generate sirnas with high activity, high stability and low toxicity. *Nucleic acids research*, 37(9):2867–2881, 2009.

[119] Mohammed Amarzguioui, Torgeir Holen, Eshrat Babaie, and Hans Prydz. Tolerance for mutations and chemical modifications in a sirna. *Nucleic acids research*, 31(2):589–595, 2003.

[120] Andrew M Kawasaki, Martin D Casper, Susan M Freier, Elena A Lesnik, Maryann C Zounes, Lendell L Cummins, Carolyn Gonzalez, and P Dan Cook. Uniformly modified 2'-deoxy-2'-fluoro-phosphorothioate oligonucleotides as nuclease-resistant antisense compounds with high affinity and specificity for rna targets. *Journal of medicinal chemistry*, 36(7):831–841, 1993.

[121] Jens Harborth, Sayda M Elbashir, Kim Vandenburgh, Heiko Manninga, Stephen A Scaringe, Klaus Weber, and Thomas Tuschl. Sequence, chemical, and structural variation of small interfering rnas and short hairpin rnas and the effect on mammalian gene silencing. *Antisense and Nucleic Acid Drug Development*, 13(2):83–105, 2003.

[122] Jonathan K Watts, Glen F Deleavey, and Masad J Damha. Chemically modified sirna: tools and applications. *Drug discovery today*, 13(19):842–855, 2008.

[123] Frank Czauderna, Melanie Fechtner, Sibylle Dames, HuÈseyin AyguÈn, Anke Klippel, Gijsbertus J Pronk, Klaus Giese, and JoÈrg Kaufmann. Structural variations and

stabilising modifications of synthetic sirnas in mammalian cells. *Nucleic acids research*, 31(11):2705–2716, 2003.

[124] Prasad Dande, Thazha P Prakash, Namir Sioufi, Hans Gaus, Russell Jarres, Andreas Berdeja, Eric E Swayze, Richard H Griffey, and Balkrishen Bhat. Improving rna interference in mammalian cells by 4'-thio-modified small interfering rna (sirna): effect on sirna activity and nuclease stability when used in combination with 2'-o-alkyl modifications. *Journal of medicinal chemistry*, 49(5):1624–1634, 2006.

[125] Divya Sivaraman, Payal Biswas, Lakshmi N Cella, Marylynn V Yates, and Wilfred Chen. Detecting rna viruses in living mammalian cells by fluorescence microscopy. *Trends in biotechnology*, 29(7):307–313, 2011.

[126] Aaron W Lifland, Chiara Zurla, Joana Yu, and Philip J Santangelo. Dynamics of native β-actin mrna transport in the cytoplasm. *Traffic*, 12(8):1000–1011, 2011.

[127] Anthony S Walsh, HaiFang Yin, Christoph M Erben, Matthew JA Wood, and Andrew J Turberfield. Dna cage delivery to mammalian cells. *ACS nano*, 5(7):5427–5432, 2011.

[128] Carlos Ernesto Castro, Fabian Kilchherr, Do-Nyun Kim, Enrique Lin Shiao, Tobias Wauer, Philipp Wortmann, Mark Bathe, and Hendrik Dietz. A primer to scaffolded dna origami. *Nature methods*, 8(3):221–229, 2011.

[129] Yang Song, Xuemei Zhang, Lingyan Huang, Mark A Behlke, and Andrew Tsourkas. Real-time imaging of single engineered rna transcripts in living cells using ratiometric bimolecular beacons. *JoVE (Journal of Visualized Experiments)*, (90):e51544–e51544, 2014.