Applications of dynamic nucleic acid nanotechnology in closed-loop genetic circuits and detection of viral pathogens

Thesis by
Jining Huang

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
Defended December 2, 2022

To my mother.

ACKNOWLEDGEMENTS

The journey through graduate research is sometimes a long, ardous and isolating one. Nature can be a cruel mistress when it comes to revealing her secrets to aspiring graduate students. There were periods where nothing seemed to work, or worse, when projects seemed to go backwards. Given these pressures throughout the process, it was imperative for the aspiring graduate student to surround themselves with a supportive environment. I'd like to take this section to give thanks to those supportive people who made that environment possible.

First and foremost, I would like to thank my advisor Niles Pierce. Niles has always shown me both the courage to plunge into the complete unknown, despite sometimes what may seem like insurmountable odds, and yet at the same time, being rigorous enough to turn a concept on paper into a useful and presentable result. On a personal note, I've always admired his leadership style, especially in tougher times. I will not soon forget the period where we were taking phone calls every 2 weeks with our funding agency, or the initial stages of the COVID-19 pandemic.

I am also grateful to my committee, for their support and encouragement both inside and outside of committee meetings. As with most students, I was quite intimidated by my committee at first. However, my committee has always me feel good about the work I was doing and treated me well during our earnest discussion of the work. I would like to thank my committee chair Erik Winfree for always helpful and positive discussions, and for leading me to think more quantitatively about biological systems. I would like to thank Michael Elowitz for being very supportive of my work, and giving excellent suggestions on both the science as well as communication techniques in committee meetings and the DARPA meetings. I would like to thank Richard Murray for everything from taking the reins of the DARPA project, giving useful and actionable advice both in and out of the scientific arena, and being a role model in organizational clarity.

I have been fortunate in having colleagues that are some of the brightest young minds and most supportive people in science of our generation. I'm deeply grateful to both Lisa Hochrein and Maayan Schwarzkopf for their scientific advice on the wet-lab projects, but more importantly their sage wisdom on graduate school and life, and their uncanny ability to make me feel better after talking to them, especially when my projects have hit a brick wall. I'm also thankful to my collaborators: Mikhail

we would see each other on a nearly daily basis, we would always chat about careers, school, and life, and because we came from drastically different worlds, I would always leave these conversations with a fresh perspective on the human experience. I'm also grateful to Alex Jong. Alex and I met when we were both in the gorgon pit of graduate work, when very little was working and we were both petrified to be gazing down our respective impending deadlines. We often ate and hung out together, and vented to each other about work and life. Alex also came from a very different background than me, and aside from the comforting conversations, Alex would often leave me thinking about my own life and motivations. After Alex left Caltech, he inducted me into a guild of misfits and ne'er-do-wells, and I met Audrey, Jeff, Sam, Patrick, Seth and others, who provided a great deal of encouragement during the thesis writing process.

I'd also like to thank Edward Panning and Constantin Busiouc, who were teachers from high school that oversaw our robotics team. These teachers made a meteoric impact during my formative years. I initially honed my skills as an engineer in these years. Unlike all the school work up to that point, I was introduced to an environment where there was little precedence in what I was doing, and I had to fail over, and over, and over again to reach success. I truly believe doing original research would not be possible without what I learned during this period. I would also like to mention the life-long friends I've made and kept to this day from that period: Brett, Frank, Rishi, Zack, David, and Alistair.

Finally I'd like to thank family. I'm grateful to my extended family for their continual support, especially Shuyue. I would like to thank my father Gang Huang, and especially my late mother Pei Liu for cultivating my passion for science from a very young age. I would also like to thank them for moving to Canada, before I can even grasp the difficulties and hardships that would bring. However, due to this move, I was able to thrive in a more holistic education system, culminating in this thesis as the completion of my formal education.

ABSTRACT

Nucleic acid nanotechnologies have provided a platform where biologically relevant molecules can be engineered to perform programmable functions. Relative to proteins, complex nucleic acid-based systems can be designed more readily due to the countable nature of base-pairing interactions and readily available physical models. These features of nucleic acids enable us to design novel interaction pathways and functions by providing well-behaved molecular mechanisms. Two examples of these mechanisms are the conditional guide RNA (cgRNA) and the hybridization chain reaction (HCR). A cgRNA is a conditional programmable regulator where an expressed RNA trigger can conditionally turn on or off transcriptional regulation. HCR is a molecular mechanism for *in vitro* and *in situ* amplification of signals to spatially identify proteins, RNA, or DNA in a sample. This thesis will first demonstrate the use of these nucleic acid molecular mechanisms in closed-loop genetic circuits and infectious disease testing using cgRNAs and HCR, respectively, then provide updated tools for the nucleic acid design community to exploit the programmable nature of nucleic acids.

We begin by demonstrating the use of conditional programmable cgRNAs in closed-loop genetic circuits. Synthetic genetic circuits allow scientists to engineer arbitrary molecular interactions in living organisms. Feedback circuits in particular are recurrently found in nature and enable useful functionalities. However, protein components of genetic circuits cannot be designed scalably, are often mined from preexisting genomes, and present difficulties in being biologically orthogonal to themselves or the host organism. We are motivated to address these limitations by using orthogonal nucleic acid circuits created *de novo*. One potential component of these circuits are conditional guide RNAs (cgRNAs). cgRNAs are switchable transcriptional regulators, and this allows gene expression to be modulated through the expression of a small RNA trigger. Here we assess cgRNAs as a component for feedback genetic circuits. As an initial demonstration of cgRNA synthetic circuits, we built and validated a simple threshold circuit and demonstrated its orthogonality and scalability by showing independent circuit functions of two switches in a single cell. We also created a larger toggle switch that is made from the same components as the previous switches. These experiments show the orthogonality and feedback capabilities of cgRNAs will position them as a composable component for scalable synthetic biology.

We then used the hybridization chain reaction mechanism to develop an adaptable and sensitive test for the detection of SARS-CoV-2. The lateral flow assay format enables rapid, instrument-free, at-home testing for SARS-CoV-2. Due to the absence of signal amplification, this simplicity comes at a cost in sensitivity. Here, we enhance sensitivity by developing an amplified lateral flow assay that incorporates isothermal, enzyme-free signal amplification based on the mechanism of hybridization chain reaction (HCR). The simplicity of the user experience after the test begins is maintained by using a disposable 3-channel lateral flow device to automatically deliver reagents to the test region in three successive stages without user interaction. Prior to starting the test, a 15-minute heat step is required. Detecting gamma-irradiated SARS-CoV-2 virions in an extraction buffer, the current amplified HCR lateral flow assay achieves a limit of detection of 200 copies/μL using nucleic acid probes to target the SARS-CoV-2 RNA genome. By comparison, five commercial unamplified lateral flow assays that use proprietary antibodies to target the viral nucleocapsid protein exhibit limits of detection of 500 copies/μL, 1000 copies/μL, 2000 copies/μL, 2000 copies/μL, and 20,000 copies/μL. By swapping out nucleic acid probes to target different pathogens, amplified HCR lateral flow assays offer a platform for adaptable and sensitive at-home testing for emergent diseases.

Components for the previous two projects are designed and analyzed with NUPACK. NUPACK is a growing software suite for the analysis and design of nucleic acid structures, devices, and systems serving the needs of researchers in the fields of nucleic acid nanotechnology, molecular programming, synthetic biology, and across the life sciences. NUPACK algorithms are unique in treating complex and test tube ensembles containing arbitrary numbers of interacting strand species, providing crucial tools for capturing concentration effects essential to analyzing and designing the intermolecular interactions that are a hallmark of these fields. The all-new NUPACK web app (nupack.org) has been re-architected for the cloud, leveraging a cluster that scales dynamically in response to user demand to enable rapid job submission and result inspection even at times of peak user demand. The web app exploits the all-new NUPACK 4 scientific code base as its backend, offering enhanced physical models (coaxial and dangle stacking sub-ensembles), dramatic speedups (20–120× for test tube analysis), and increased scalability for large complexes.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Mark E. Fornace, Jining Huang, Cody T. Newman, Nicholas J. Porub-sky, Marshall B. Pierce, and Niles A. Pierce. NUPACK: Analysis and design of nucleic acid structures, devices, and systems, November 2022. URL `https://chemrxiv.org/engage/chemrxiv/article-details/636c7089b588507d0045f283`.
J.H. participated in the conception of the project, designed the architecture with M.E.F., participated in user interface design with M.E.F., C.T.N., and N.A.P, built the equipment and infrastructure, wrote the control plane, and participated in preparation of the manuscript.

[2] Samuel J. Schulte, Jining Huang, and Niles A. Pierce. HCR lateral flow assays for amplified instrument-free at-home SARS-CoV-2 testing, September 2022. URL `https://www.biorxiv.org/content/10.1101/2022.09.18.508442v1`.
J.H. participated in the conception of the project, designed, manufactured, and optimized the RNA detection tests, prepared publication data for RNA tests, and participated in preparation of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are nature's medium of information storage and transfer. DNA is often thought of as a double helix structure and stores genetic information. RNA is often thought of as messengers that are transcribed from DNA and facilitate translation of proteins. While these are the most common occurrences of nucleic acids in nature, nucleic acids, especially RNA, often form mechanistic components necessary for cellular function. RNA, in the form of tRNAs and rRNAs, is as much the machinery of translation as it is information to be stored and disseminated. RNAs can also form structures, such as hairpins, that can promote binding of the machinery of translation. The function of these functional nucleic acids are determined by their secondary and tertiary structure, which is in turn, determined by their sequence. The secondary structure of a complex of nucleic acids is defined by a list of base pairing interactions, and the tertiary structure is the three-dimensional arrangement of the nucleic acid molecule as determined by those base pairing interactions.

The parameters for nucleic acid hybridization for RNA [2] and DNA [3] have been established since the 1990s. A coarse grained physical model for nucleic acid hybridization exists due to the exclusionary nature of base pairing. This model allowed researchers to create algorithms to predict structural and base pairing behavior given the sequence of a nucleic acid strand [4]. This work was further expanded upon by predicting the concentration of secondary structures given a tube of interacting strands [5]. Nucleic acid secondary structures in a complex of multiple strands could be designed by exploiting the thermodynamic properties of sequences of nucleic acids calculated *in silico* [6]. Design for specific test tube concentrations of secondary structures [7] and entire pathways of orthogonal interaction networks [8] were later developed.

The rational design of nucleic acid systems enables researchers to create structures based on sequence-guided self-assembly, such as DNA tiles and origamis [9, 10]. Dynamic interactions between nucleic acid strands are also exploited to form reaction pathways and networks [11, 12] *in vitro*. In the cellular setting, rational design of nucleic acids have enabled transcriptional [1, 13] and translational [14] regulation

based on the conditional presence of nucleic-acid triggers. In imaging, nucleic acids have been used to amplify multiplexed signals *in situ* to observe mRNA, DNA, and proteins spatially in various organisms [15–21].

This thesis describe two projects that stem from the fruits of nucleic acid rational design and another to further progress the tools of nucleic acid rational design. In the first project, we created orthogonal and versatile synthetic feedback circuits using small conditional nucleic-acid based transcriptional regulators. In the second project, we created a cheap, robust, and scalable molecular diagnostic to detect SARS-CoV-2 RNA using a nucleic acid polymerization reaction. The last project describes new developments in NUPACK, a tool used by researchers world wide to design and analyze nucleic acid systems.

1.1 Nucleic acids in synthetic genetic circuits

Genetic circuits exist in nature to serve a variety of functions. A classic example is the regulation of lactose metabolism [22], where a repressor is deactivated in the presence of lactose, allowing transcription of metabolic enzymes for lactose. Additional genetic regulation circuits in bacteria include metabolism of arabinose [23], resistance to tetracycline [24], biofilm formation through quorum sensing [25], and chemotaxis [26]. These initial circuits spurred researchers to seek motifs in natural networks [27], and many more natural regulatory networks have been elucidated in both eukaryotic and bacterial cells. These include the metabolism of galactose in yeast [28], homeostasis of calcium in cows [29], and defense against foreign invaders for various species [30–32]. The regulation of development and growth also utilize genetic circuits. Examples include differentiation of wings [33] and embryos [34] in *D. melanogaster*, and regulation of plant growth [35].

Taking inspiration from nature, researchers have created novel synthetic genetic circuits by commandeering existing protein components. Various biomolecular devices were created in *E. coli*, beginning with the repressilator [36] and toggle switch [37], which took advantage of orthogonal systems of repressor proteins. Further functionality such as logic gates [38, 39], event counters [40], pulse generators [41], and comparators [42] were created once researchers gathered more activating protein components. Devices with permanent digital memory [43] and temporally controlled logic gates [44] were created once circuit engineers added integrases and excisionases to their repertoire of tools. The identification of multi-cell communications systems such as quorum sensing enabled researchers to build multi-cellular

circuits such as cells that pattern themselves spatially [45]. Further studies into the kinetics and dynamics of interaction between components has led to better behaving repressilators [46]. Recent advances in post-transcriptional components enabled researchers to connect proteases in various ways to create more composable circuits [47].

A particularly interesting class of circuits are closed-loop circuits and feedback circuits due to their many functional features [48]. The simplest closed-loop feedback circuit is transcriptional autoregulation. This motif has been demonstrated in bacteria [49], and it enables faster expression rise time than the open-loop alternative. The repressilator and toggle switch are both closed loop circuits, establishing capabilities for oscillations and bistability. The incoherent feedforward loop is a closed loop circuit, although it does not contain feedback. This circuit acts as a band-pass filter [47], and it has been theorized to act as a fold change detector [50]. Feedback has also been used as a motif to control cell behavior. For example, feedback was used to regulate total population of bacteria [51] and track intracellular concentrations of particular input molecules [52].

While synthetic biology has accomplished a great deal in the past two decades, there still are challenges in scaling synthetic circuits. Demonstrations to date have shown circuits with around five components or less. Researchers are typically satisfied with a single functional circuit created and validated in a single, often specialized, cell strain. Multiples of the same type of circuit often cannot operate in the same cell due to shared protein components leading to deleterious crosstalk [53]. To further the field, there need to exist more, better characterized, and well-behaved orthogonal components [53–58] to allow researchers to use arbitrary numbers of identical circuit motifs in the same cell. For example, a feedback control circuit requires a fast controller [59]. A sequestration motif is a natural fit for this role, and while there are various implementations of protein sequestration [42, 52, 60], it would be difficult to utilize all these systems in the same cell and impossible to utilize multiples of the same system in the same cell. This issue of scale is not lost on researchers. There are attempts to use orthogonal protein proteases from mammalian genomes [47] and to use directed evolution to create larger libraries of orthogonal regulators [61].

With nucleic acid design motifs, such as toehold mediated strand displacement, the advancements in rational design has enabled researchers to create large interacting networks of DNAs *in vitro*. The computing power of interacting DNA strands is

demonstrated by evaluating the square root of an input [62]. Attempts were recently made to integrate nucleic acid circuitry into living cells to perform therapeutic and probing roles [11]. Current developments in the field of nucleic acid regulation include orthogonal sets of riboswitches [14], which controls translation, and conditional termination [13], which controls transcription. The reaction motif and secondary structure design for these regulators can be the same; however, the unique sequences for each orthogonal system are designed to explicitly prohibit crosstalk. There are various hybrid feedback circuits [51, 63, 64] which take advantage of these components along with protein components. However, these mechanisms have been limited to bacterial cells and there is little work done to bring fully nucleic acid circuits in bacteria.

Recently, a programmable gene regulation mechanism was engineered from a bacterial immune system known as CRISPR (clustered regularly interspaced short palindromic repeats). CRISPR uses a 20-nucleotide spacer gathered from virus samples to direct a Cas9 (CRISPR associated protein 9) protein to cut any double stranded sequence bearing the reverse complement [32]. Researchers commandeered this system for gene editing [65, 66] and gene regulation. dCas9 (dead Cas9), with a defunct catalytic center and various attached effectors, can activate [67–69] or silence [67, 70] expression in bacterial [67, 69, 70] and mammalian [68, 70] cells directed by gRNAs (guide RNAs). Investigation into bacterial [71], yeast [72], and mammalian [73, 74] CRISPR circuits yielded multiplexed gene expression [73], CRISPR logic gates [71], and gRNA expression cascades of up to 7 regulators [72, 74]. A CRISPR feedback circuit was recently created by expressing a miRNA-mKate-gRNA fusion. The miRNA was used to degrade the entire transcript, and the loop was broken by using Csy4 to cleave the RNA between the gRNA and miRNA binding site [73]. Another group created a metabolic feedback circuit utilizing a promoter sensitive to metabolic burden to induce gRNA silencing an artificial metabolic load [75].

Although gRNAs are programmable, they are not conditional. While Cas9 [76] or gRNA [77] can be induced by small molecules, they are not dynamic as is required in gene circuits. cgRNAs (conditional guide RNAs) are gRNAs which can be turned on or off by a small trigger RNA. Researchers can design arbitrary sets of orthogonal cgRNAs which interact in predictable ways due to the programmable nature of nucleic acid interactions. In addition to orthogonality and scalability, the conditionality of cgRNAs introduce an RNA-based sequestration motif which circuit designers can take advantage of [1, 78–80].

The capabilities of cgRNAs can be summarized as follows: 1) programmable transcriptional effector, which can silence or activate expression based on the sequence identity of a 20 nucleotide spacer, 2) conditional activation/deactivation by short trigger strands providing a sequestration motif with RNA hybridization time scales of interaction as opposed to transcriptional time scales, 3) large sequence space is available for design of orthogonal cgRNAs and orthogonal effector sites, 4) insertions of sequence, such as DNA binding sites, into the 5' end of cgRNAs and triggers can be tolerated to enable dCas9 transcriptional control, and 5) cgRNA interactions with its triggers are orthogonal. The latter combined with the 3rd and 4th points allows researchers to design scalable and orthogonal systems of interacting cgRNAs and triggers and wire these interacting systems together in arbitrary ways. Chapter 2 of this thesis explores the novel features cgRNAs can bring to closed-loop genetic circuits.

1.2 Nucleic acids in molecular diagnostics

In March of 2020, the SARS-CoV-2 virus and its corresponding disease COVID-19 spread to the United States. Uncertainty loomed as lock downs were rampant and available testing for the novel pathogen was scarce. The genetic sequence of the novel coronavirus, Wuhan-Hu-1, was shared through an online database on January 5th, 2020, and the first laboratory-confirmed case outside China was announced on January 13th, 2020 [81]. It only took, at most, a little more than a week to create a laboratory test for the novel virus after the initial publication of the genomic RNA sequence. These initial laboratory tests are what's known as RT-qPCR (reverse transcription quantitive polymerase chain reaction, sometimes just PCR) tests. Briefly, a reverse-transcriptase creates the cDNA (complementary DNA) from the RNA genome of the virus, then qPCR is performed to quantify the abundance of cDNA present in the reaction. These tests take relatively little time to adapt for emergent diseases, as the only changed component for a new RNA genome is the DNA probes. DNA probes can be synthesized chemically, and have fast manufacturing turnaround. However, as the pandemic progressed, the demand for testing also grew. As the pandemic was spreading to the United States, there were shortages of PCR supplies. Exacerbating the issue, RT-qPCR requires specialized equipment, and there was also an equipment shortage. These shortages led to efforts to conserve equipment and reagents [82].

Eventually, a new form of testing, complementary to RT-qPCR, would come to the market: the rapid antigen test. This new test uses the same technology as a pregnancy

test. Briefly, on a lateral flow strip, an antibody to the SARS-CoV-2 N (nucleocapsid) protein is immobilized at the test line, and on a conjugate pad, another antibody to the SARS-CoV-2 N protein is conjugated to a visible chromophore. The sample solution travels from the sample pad, to the conjugate pad, where the antibody with chromophore binds to the N protein, and finally up to the test region, where the immobilized antibody binds to the N protein-antibody-chromophore complex and concentrates the chromophore on the test region. This concentration of chromophore creates a visible signal for the human eye if the virus is sufficiently abundant in the provided sample.

This new testing paradigm provides a rapid, inexpensive, robust and simple test. A usual antigen test only takes 10 to 15 minutes, can be stored at room temperature for an extended period of time, and can be performed by the consumer at home. However, the first variant of this new testing form was seen on the market in May [83], 4 months after the publication of the genome, and the initial implementation of the test required specialized fluorescence equipment. The first widespread rapid antigen test, that is suitable for home use, with no equipment requirements, was released on August 26 [84], a full 7 months after the initial sequencing of the virus. Another issue with the rapid antigen tests is their lack of sensitivity, leading to these tests having a high false-negative rate (e.g., 25–50% in two hospital studies)[85, 86].

Recently, newer technologies are being evaluated in academic and startup circles for use in pathogen testing in general and SARS-CoV-2 in particular. One of these approaches is using RT-LAMP (reverse transcription loop-mediated isothermal amplification) [87, 88]. RT-LAMP follows the same basic steps as RT-qPCR with several key differences. In the DNA amplification phase, special primers are used such that polymerized DNA form a hairpin loop on either end, such that additional self-primings can occur in the loop [89]. This change allows RT-LAMP to be performed isothermally. A pH indicator is added that changes the color of the entire reaction solution if DNA is synthesized. This method is simple, sensitive, and requires less sophisticated equipment, however it has some of the same drawbacks as RT-qPCR. RT-LAMP uses enzymes, which require cold storage, and often needs a heat source, which requires laboratory equipment.

Another approach is using CRISPR as a readout [90, 91]. Briefly, the viral genomic RNA is amplified, the amplified DNA is transcribed by IVT (*in vitro* transcription). The RNA transcribed is introduced to a Cas protein (Cas12a or Cas13a) with a specific target gRNA, once the Cas recognizes the binding site on the transcribed

RNA, it indiscriminately cleaves single stranded nucleic acid with two haptens on either side of the cleavage site. The cleavage or non-cleavage of the single stranded nucleic acid can be read in several ways including lateral flow. This method is not as sensitive as RT-qPCR, but it does require less sophisticated equipment. While its steps are not fully isothermal, it does not require constant thermocycling. However, the inclusion of a heating steps will require some form of laboratory equipment. Enzymes, with their cold storage requirements, are also required for the amplification, IVT, and CRISPR detection steps, leading to more cold storage requirements. As both RT-LAMP and the CRISPR-based tests are nucleic acid tests, they could in theory be adopted quickly to novel emergent pathogens.

We set out with the goal of creating an amplified instrument-free rapid test that can be deployed quickly in the event of an emergent disease, have better sensitivity than commercial rapid antigen tests, is robust to supply chain demands, is inexpensive, and is simple to use. To quickly deploy these tests in emergent diseases, we chose to detect viral RNA with DNA probes. DNA probes can be synthesized cheaply and quickly once the genomic sequence of the novel pathogen is available. However, due to the low abundance of available RNA genomes in the initial stages of infection, amplification is required. To maintain our goal of creating a robust test, we decided to avoid enzymes for amplification, as they are difficult to manufacture and require specific storage and operational conditions. We proposed using nucleic acid HCR (hybridization chain reaction) for our amplification.

HCR is used to spatially image RNA, DNA, or proteins in various organisms *in situ* [15–17, 20, 21]. HCR imaging works in two stages. First, a probe is bound to the target of interest, then amplification takes place from the probe. HCR is ideal for detection platform, as it provides signal amplification but does not require any enzymes. The entirety of the HCR amplification system only comprises DNA, a molecule that is quick to synthesize, will not deteriorate under ambient conditions, and does not require cold storage. Here, we use the HCR amplification motif to amplify signal at the test region of a lateral flow device. HCR has not been previously used on paper or thin-films. Furthermore, HCR requires multiple reagent addition, removal, and wash steps that will need to be automated to be simplistic enough to use in a point-of-care or home setting. Chapter 3 will describe the prototype of an automated device to detect SARS-CoV-2 virus using HCR amplification.

1.3 NUPACK: design and analysis of nucleic acid interactions

NUPACK is a tool to analyze and design nucleic acid structures, interactions, and pathways. The core nucleic acid components of both projects introduced in the previous sections, cgRNAs and the HCR amplification system, were designed with NUPACK. NUPACK has two main capabilities: analysis and design of nucleic acid systems. NUPACK can be provided with a set of strands in a tube and the concentration of each strand, and it will compute equilibrium physical quantities of the tube, such as concentrations of all possible nucleic acid complexes and their base pairing probabilities. NUPACK can design sequences from a tube specification of desired nucleic acid complex structures and concentrations.

Over the years, NUPACK has received several new updates to its recursions, parameters [92], and design features [7, 8], such as weights and constraints [8]. Concurrently, the NUPACK user base has grown continually. In the past year there are more than 3.5 million page views from 180,000 unique visits by researchers, spending a cumulative 2.7 million minutes on the web application. However, the current NUPACK web application still uses a code base written in 2007, and the advanced age of the hardware and software is responsible for consistently full queues and slow downs.

To address these usability issues and implement the latest features in the scientific code base, NUPACK needed to be rearchitected and rebuilt from the ground up. The new NUPACK web app will incorporate both the latest developments in the scientific backend as well as a scalable cloud-based compute infrastructure to enable the larger demand of the nucleic acid design community. Chapter 5 of this thesis will describe the new NUPACK web application, the inputs, computations, and results for both the analysis and design features of NUPACK. It will also describe and characterize the performance of the new infrastructure and computational backend written from the ground up.

1.4 Thesis overview

In this thesis I will describe each of the projects introduced above in detail. Chapter 2 will cover applications of small conditional nucleic-acid based cgRNAs in creating scalable, orthogonal, and versatile closed-loop genetic circuits. Chapter 3 will cover applications of HCR nucleic-acid signal amplification on a strip-based lateral flow platform to detect SARS-CoV-2 virus. Chapter 4 will cover NUPACK cloud, rearchitected from the ground up using the latest technologies to incorporate a new

burstable cluster, user interface, and feature updates to enable the next generation of nucleic-acid design community to create large and complex nucleic-acid systems.

C h a p t e r 2

USING CGRNAS FOR SCALABLE SYNTHETIC FEEDBACK CIRCUITS

Natural gene regulatory networks utilize feedback and large numbers of nodes to perform functions, such as development [93–96]. The misregulation of these feedback networks can lead to disease [97]. Synthetic biology takes inspiration from nature and commandeers its components to create functions *de novo*. Synthetic biologists are often motivated to create synthetic genetic circuits with a wide variety of capabilities utilizing the feedback motif, such as oscillators [36, 46], toggle switches [37], and molecular trackers [52]. These genetic circuits typically use two modes of feedback, 1) transcriptional feedback using expressed transcription factors to regulate other transcription factors, or 2) sequestration feedback using two separately expressed components that bind to activate or deactivate function. While there is a large breadth of functional modules in synthetic biology, there still needs to be a focus towards the scaling of these modules [53–58] to approach mimicking the large network of genetic circuits in nature.

While nature can use the evolutionary process to create various orthogonal pathways of interaction using mostly protein-based mechanisms, we cannot yet do so synthetically. The protein parts that most of synthetic biology is built on were mined from existing genomes in nature. Multiples of the same type of circuits cannot operate in the same cell, as the same proteins will cause deleterious crosstalk between the two circuits. Nucleic acid-based circuits, on the other hand, can be designed *de novo*, enabling them to be orthogonal to themselves and other molecules in the pathway of interaction.

Concurrent to the development of synthetic biology, the field of nucleic acid design has also greatly advanced. A physical model exists for base pairing interactions [4] from the countable nature of nucleic acid base pairing combined with experimentally determined energy parameters [2, 3]. Using this physical model, advanced nucleic acid interactions pathway designers can design nucleic acid sequences from specified nucleic acid interactions [8]. The advent of nucleic acid design led to many instances of useful nucleic acid circuits *in vitro* [11, 12, 62, 98, 99], and fewer instances of nucleic acid regulatory components in bacteria [1, 13, 14].

Programmable regulators for use in synthetic biology have been created with both with proteins [47, 61] and, with the advent of CRISPR, nucleic acids [71–74, 100] in living cells. On the protein front, orthogonal regulators can be mined from nature, and a variety of efforts have borne fruit to create larger orthogonal networks of protein circuitry by exploiting integrases [101, 102], proteases [47], and other protein components [103], but due to the difficulty of rational protein design, an arbitrary number of them cannot be created *de novo*. On the nucleic acid front, orthogonality can be achieved using a well established design process, but the libraries of functional modules currently lack behind protein circuits. A large number of gRNAs [72] and gRNA feedback [73, 100] were demonstrated in *E. coli*, but not together.
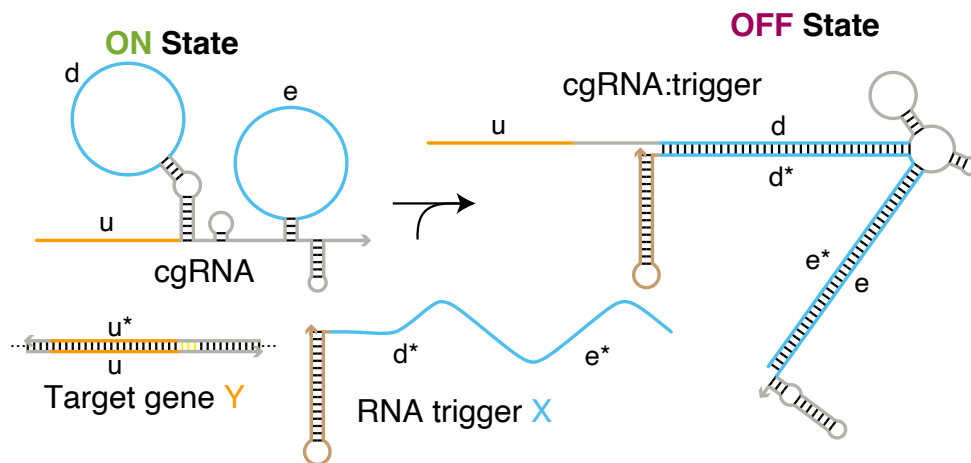
Figure 2.1: cgRNA schematic [1]: the constitutively active cgRNA is inactivated by hybridization of RNA trigger X.
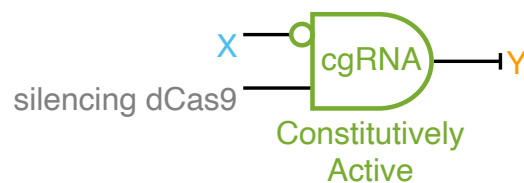
Figure 2.2: cgRNA ON→OFF logic using silencing dCas9 [1]. Conditional logic: if not X then not Y.

Towards the goal of achieving scalable synthetic circuits, the Pierce lab created allosteric cgRNAs (conditional guide RNAs) [1, 104]. cgRNAs are conditional programmable regulators, built on top of programmable guide RNAs. A cgRNA

can be activated or deactivated in the presence of a small expressed RNA trigger via structural transduction (Figure 2.1). The guide RNA controls the genomic target of regulation, while the conditional trigger controls the scope of the regulation, and the variant of Cas9 protein used controls the mode of regulation (Figure 2.2). cgRNA activity can be dynamically and allosterically regulated by a small expressed trigger. This is a substitute for the sequestration of ligand and receptor in proteins. These properties allow cgRNAs to satisfy both transcriptional and sequestration modes of feedback circuits. Circuits built with cgRNAs do not need protein circuit components (other than globally expressed Cas9) because cgRNAs are also purely transcriptional, whereas other cellular nucleic regulatory systems rely on translation.

The purpose of this work is to demonstrate cgRNAs as a useful component for closed-loop synthetic genetic circuits. We intend to do this with three demonstrations: First, we built a simple feedback circuit implemented using cgRNAs taking advantage of both the transcriptional and sequestration modes of feedback. Then we scaled up that same circuit motif multiple times in the same cell. Finally, we prototyped a larger implementation of a cgRNA circuits using the exact same components. Through these experiments, we demonstrate cgRNAs can be foundational elements for complex and interconnected circuits.

To demonstrate a simple feedback circuit, we created a thresholding switch as a proof of concept. This switch utilizes both the transcriptional regulation from CRISPR as well as regulation based on RNA hybridization. The performance of the thresholding switch was observed when feedback was disrupted when either the transcriptional or hybridization regulation was disrupted. To demonstrate scaling of this feedback circuit, the same switch can be implemented to form two working circuits that operate independently in the same cell. In this case, we use two different orthogonal cgRNAs to implement two of the same circuit schematic in the same cell. To demonstrate the possibility of larger and interconnected circuits using the same base components, we show a toggle switch where every component of the switch can be transcribed by the cell using the same circuit components as the thresholding switch.

2.1 A simple thresholding circuit

We first built a simple thresholding feedback circuit using cgRNA components as a proof of concept for incorporating cgRNAs in synthetic genetic circuits. Feedback

circuit design and a scale-up study demonstrates the advantage of hybridization and orthogonality, respectively.

Feedback circuits generally use a sequestering component to affect the feedback on a faster timescale than the rest of the circuit. One of the major benefits of cgRNAs is the dynamics of RNA hybridization [1]. We conducted a study to create a simple feedback cgRNA circuit to probe the synthetic biology applications of cgRNAs. The circuit incorporates closed-loop feedback and has an observable shift in the transfer function of the circuit between open-loop circuit controls and the closed-loop feedback circuits.

Figure 2.3: Schematic for single thresholding switch: IPTG induces cgRNA that transcriptionally silences trigger, which in turn hybridizes to cgRNA disabling it. Open-loop controls by disrupting transcriptional inhibition (lavender scissors), which provides upper bound control; and inactivation by hybridization (yellow scissors), which provides lower bound control.

In this circuit architecture (Figure 2.3), a small molecule inducer (IPTG) induces a cgRNA and the cgRNA inhibits a short trigger via a dCas9 effector. In the closed-loop setting, the trigger than inactivates the original cgRNA by hybridization. The output of the circuit is the abundance of active cgRNA, this is read out by using the cgRNA to inhibit transcription of a constitutively active fluorescent protein. The feedback loop can be opened in 2 ways: 1) removing the hybridization between the trigger and cgRNA, which would leave the cgRNA always maximally active and in turn, fluorescence signal would always be low, creating the lower bound control, and 2) removing the transcriptional inhibition between the cgRNA and trigger, which would lead to the cgRNA being maximally repressed due to high trigger production leading to a high fluorescence signal, creating the upper bound control.

The behavior of the circuit was validated in a modified MG1655 strain of *E. coli* (Ec001, a gift from Dr. Stanley Qi) [70]. The strain contains a constitutively expressed RFP, GFP, and Kanamycin resistance gene in the nfsA locus. The RFP and GFP genes are used as outputs for circuits on transformed plasmids.

Figure 2.4: Plasmid layouts and labels for orthogonal thresholding circuits. Crooked arrows indicate promoters, forward triangles indicate cgRNA targeting sites, backward triangles indicate dCas9 binding sites. Black blocks indicate cgRNAs and triggers.

Plasmid	FP	cgRNA	Trigger 1	Trigger 2	Function
P4	GFP	D	D	C	Closed Loop
P5	GFP	D	C	C	Open by hybridization
P6	GFP	D	C	D	Open by transcription

Table 2.1: Descriptions of thresholding circuit plasmids using GFP as an output. Circuit plasmid using layout in Figure 2.4.

The circuit is composed of cgRNAs and triggers on plasmids targeting GFP as an output (Figure 2.4). Three transcriptional units are placed in sequential order to avoid unexpected enhancement or inhibition of expression due to plasmid context effects [105]. The first transcript is a cgRNA targeting GFP under a pLac promoter, inducible by IPTG. The second transcript is a NGG–GFP cgRNA–trigger fusion, allowing this transcript to be transcriptionally affected by the cgRNA and act as the trigger. The third transcript is an additional orthogonal trigger sequence. Two other plasmids were created to be the open-loop control circuits. They have different triggers on the second and third transcript. In the open-loop circuit with a broken hybridization arrow, the second and third transcript both have a non-cognate trigger. In this case, there is no cognate trigger. In the open-loop circuit with a broken transcription arrow, the second transcript has a non-cognate trigger and the third transcript has a cognate trigger. In this case, the cgRNA would have no effect on the production of trigger. The number of transcriptional units and dCas9 binding sites have been kept constant to minimize the effects of varying dCas9 occupancy and metabolic load of the circuits.

Figure 2.5: Data for thresholding circuits using GFP as output. The circuit that is open-loop by transcription sets the upper bound, the circuit that is open-loop by hybridization sets the lower bound, the closed loop circuit is thresholding between these bounds. Data are mean ± standard deviation of N=3 replicates. Each replicate is normalized by the fluorescence/OD600 value of the closed-loop circuit under the IPTG=0μM condition. Refer to Section A.4 for detailed replicates.

The transfer function for the three circuits are as expected (Figure 2.5). The open-loop control with a broken hybridization arrow by using a non-cognate trigger is the lower bound. The fluorescence is low for any induction conditions since the cgRNA is always maximally active due to the lack of inhibition by trigger. The open-loop control with a broken transcriptional arrow by disassociating the dCas9 binding site with the trigger is the upper bound. The fluorescence is high for any induction conditions because the cognate trigger is maximally constitutively expressed with no dCas9 transcriptional inhibition. The data shows that across the transfer function, the open-loop circuits form the upper and lower bounds and the closed-loop circuit behaves as a threshold switch between these bounds. This is the first demonstration of cgRNA circuits and CRISPR feedback circuits involving cgRNAs.

2.2 Independently functioning circuits in the same cell

We next explored using orthogonal cgRNAs to bring the scalability of nucleic acid dynamic nanotechnologies to living organisms. To demonstrate scalability, we first need to demonstrate the orthogonal nature of multiple cgRNAs to create another set of thresholding switches with the same circuit schematic and plasmid layout, but

different and orthogonal sequence design targeting RFP instead of GFP. In order to switch the second set of circuits to silence RFP, the cgRNA targeting region was changed to a RFP mRNA sequence. The cgRNA efficacy was reduced and the kinetics of the circuit was altered with this change.
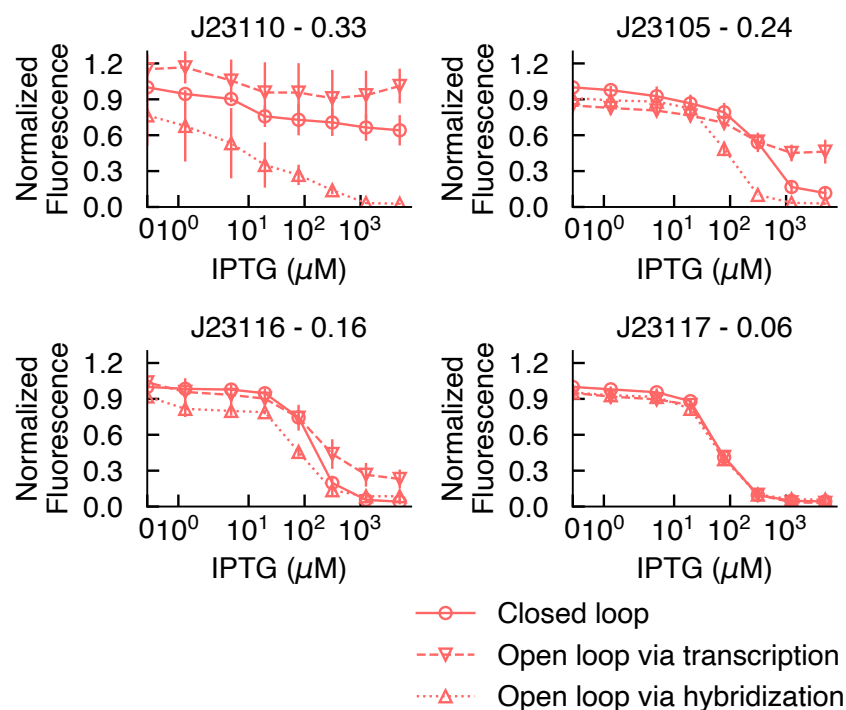
Figure 2.6: Data for thresholding RFP circuit optimization. As promoter strengths decrease, the two open-loop circuits collapse the usable range of induction. As promoter strength increases, the closed-loop thresholding circuit thresholds at above the available range of induction. Data are mean ± standard deviation of N=3 replicates. Each replicate is normalized by the fluorescence/OD600 value of the closed loop circuit under the IPTG=0μM condition. Refer to Section A.4 for detailed replicates.

Plasmid	FP	cgRNA	Trigger 1	Trigger 2	Function
P1	RFP	B	B	C	Closed Loop
P2	RFP	B	C	C	Open by hybridization
P3	RFP	B	C	B	Open by transcription

Table 2.2: Descriptions of thresholding circuit plasmids using RFP as an output. Circuit plasmid using layout in Figure 2.4.
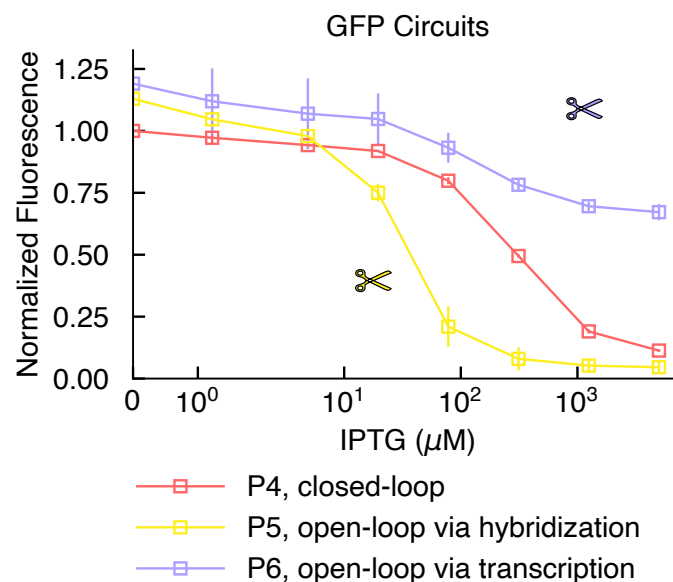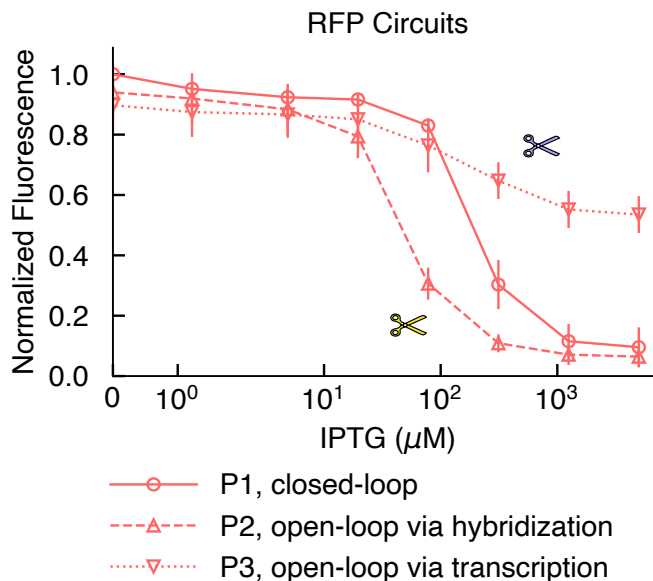
Figure 2.7: Data for thresholding circuits using RFP as output. The circuit that is open-loop by transcription sets the upper bound, the circuit that is open-loop by hybridization sets the lower bound, the closed loop circuit is thresholding between these bounds. Data are mean ± standard deviation of N=3 replicates. Each replicate is normalized by the fluorescence/OD600 value of the closed-loop circuit under the IPTG=0μM condition. Refer to Section A.4 for detailed replicates.

A series of optimizations were performed for the RFP circuit (Figure 2.6). The promoter on the second and third transcriptional units were altered for all open- and closed-loop circuits with various Anderson promoter [106] strengths (refer to Section A.4 for experimental details). The Anderson promoters tested has a quantified relative strength of 6% to 33%. Experiments show that as the promoter strength increases, the open-loop controls exhibit a larger dynamic range and the closed-loop circuit exhibit a distinctively different transfer function between the open-loop controls. As promoter strength is increased too high, the closed-loop transfer curve converges with the open-loop with broken transcriptional control as the thresholding region falls outside of the induction range. The promoter at a relative strength of 24% is found to have a larger than noise dynamic range between the open-loop controls as well as a well behaved closed-loop circuit. The circuit was built to the same plasmid specifications as the GFP circuit with a different origin and resistance marker to allow the two circuits to co-exist in the same cell (Figure 2.8 and Table 2.2). The RFP circuit was observed to behave similarly to the GFP circuit (Figure 2.7). The upper-bound, lower-bound and thresholding behavior corresponded with the open-loop by transcription, open-loop by hybridization, and

closed-loop circuits, respectively. The orthogonality of modules shown in these experiments is a solid foundational step towards building more complex cgRNA circuits.
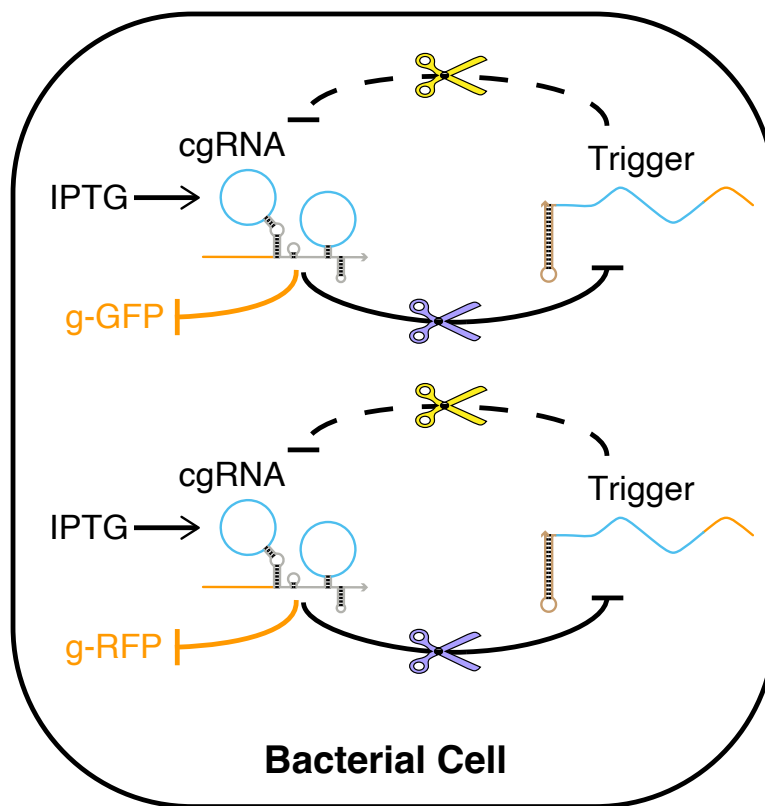
Figure 2.8: Schematic for two threshold switches in a single cell, both switches are activated by IPTG, however the switches are orthogonal to one another via orthogonal sequences.

Once a well behaving RFP circuit was observed, both orthogonal circuits were transformed into the same cell (Figure 2.8). The circuits described previously are transformed in a pairwise fashion, where every circuit plasmid from the RFP system was paired with every circuit plasmid from the GFP system. There are 9 strains total containing all possible pairwise combinations of both systems. In all strains with plasmids P4, P5, and P6, the GFP fluorescence thresholds, is a lower-bound, or is an upper-bound, respectively, regardless of whether plasmids P1, P2, or P3 is also in the cell. This can be observed in Figure 2.9 as all identically colored lines are clustered together in the GFP circuits plot. Conversely, in all strains with plasmids P1, P2, and P3, the RFP fluorescence thresholds, is a lower-bound, or is an upper-bound, respectively, regardless of whether plasmids P4, P5, or P6 is also

in the cell. This can be observed in Figure 2.9 as all lines with identical styles and markers are clustered together in the RFP plot. This verifies that the GFP circuit in a cell does not interfere with a RFP circuit and vice-versa, demonstrating two independently operating circuits using the same motif operating in the same cell.
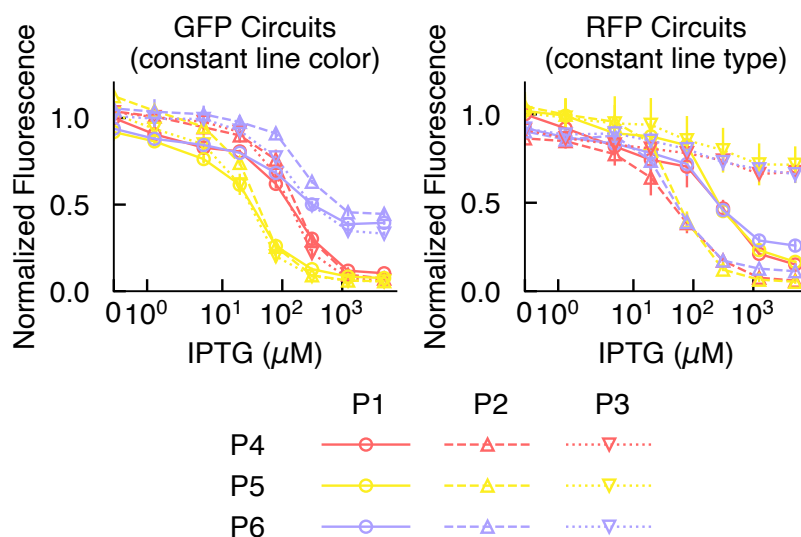
Figure 2.9: Data for two independent orthogonal thresholding circuits in the same cell using GFP and RFP as output. Only the GFP plasmid affects the GFP output, and only the RFP plasmid affects the RFP output. Data are mean ± standard deviation of N=3 replicates. Each replicate is normalized by the fluorescence/OD600 value of the strain with both closed-loop circuits under the IPTG=0μM condition. Refer to Section A.4 for detailed replicates.

2.3 Versatility of cgRNAs to build larger circuits
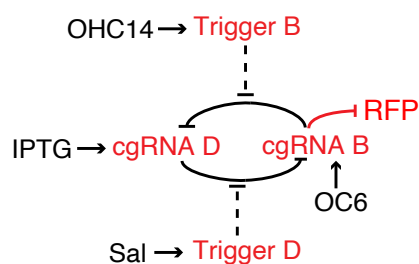
Figure 2.10: Toggle switch schematic. cgRNAs silences each other to create bistable switch, triggers induced by small molecules disrupt cgRNA activity to actuate switch.

In order to demonstrate the programmability of cgRNA technology, we created a larger circuit with the same cgRNA and trigger components as our independent

smaller circuits. We chose to build a toggle switch circuit with RNA-based inputs and outputs, using triggers and cgRNAs in place of small molecules and protein repressors (Figure 2.10).

This circuit was built in a modified Marionette-WT strain of *E. coli* (a gift from Dr. Chris Voigt) [61]. The strain contains constitutively expressed repressors for 12 different small molecule inducers. We used 4 of these inducers for our cgRNA and trigger components. Some of these are used as circuit inputs (i.e., the trigger inducers) while others are used to parameterize the circuit (i.e., the cgRNA inducers). We further modified the strain by genomically incorporating a constitutively expressed mRFP1 fluorescent protein as the readout.
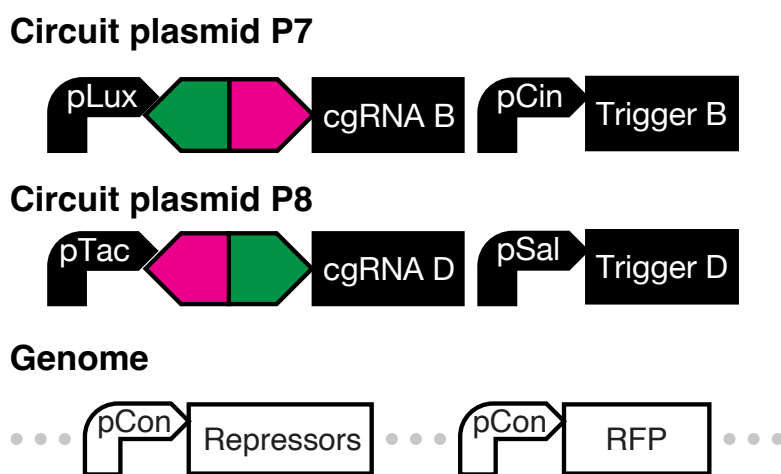
Figure 2.11: Plasmid layouts for toggle switch circuits and genomically incorporated elements. Crooked arrows indicate promoters, forward triangles indicate cgRNA targeting sites, backward triangles indicate dCas9 binding sites. Red and green arrows indicate sites for RFP and GFP, respectively. Black blocks indicate cgRNA and trigger.

The circuit is composed of cgRNAs and triggers on two plasmids (Figure 2.11). Each plasmid consists of two transcription units of one cgRNA and one trigger of the same system. All transcriptional units use a different repressor system in the Marionette strain. We deliberately made the decision to place all dCas9 binding sites on separate plasmids to minimize any physical effects of dCas9-plasmid binding.

We first determined the parameters for cgRNA induction. An experiment was ran with a dilution matrix of cgRNA B and cgRNA D under the condition that either trigger B or trigger D was expressed. We aimed to see a cgRNA expression condition such that when trigger D is expressed, RFP signal is low, and when trigger

Toggle circuit time course behavior
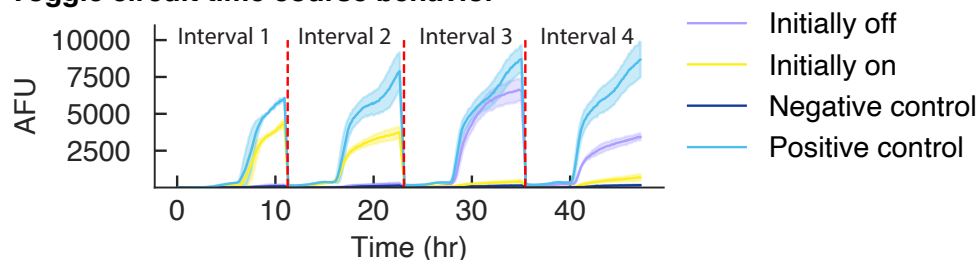
Toggle circuit end-of-interval flow cytometry data
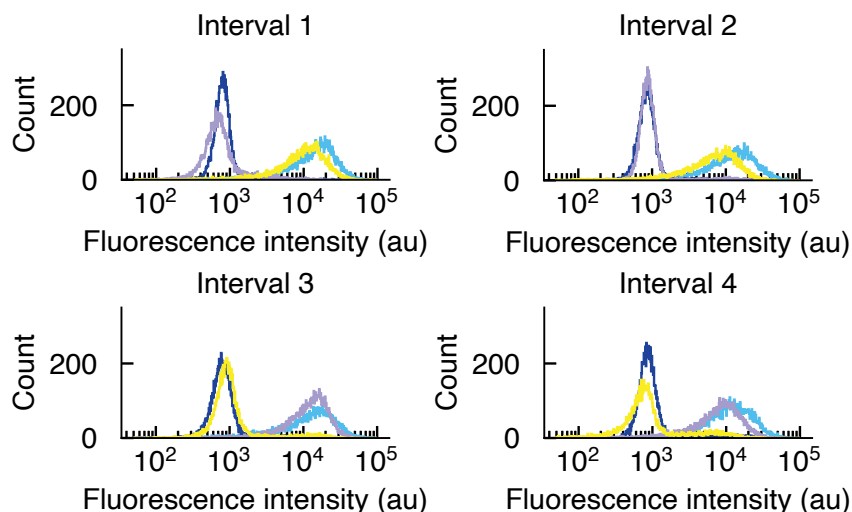
Figure 2.12: Validation of RFP toggle switch circuits using time course fluorescence measurements and flow cytometry. Time course fluorescence measurements are mean ± standard deviation for N=3 replicates. One out of three flow cytometry replicates are shown. Refer to Section A.4 for detailed replicates. For experiment lines (initially on and initially off) cgRNA B and cgRNA D are induced with 100nM OC6 and 100μM IPTG, respectively, for all intervals. Trigger B and trigger D are induced with 10μM OHC14 and 100μM Sal, respectively, for intervals 1 and 3. After removing trigger expression via dilution of trigger inducers, cells maintain their fluorescence as the previous interval with trigger. Negative fluorescence and positive fluorescence were induced with only 1μM OC6 and 1mM IPTG, respectively, for all intervals.

B is expressed, RFP signal is high. We found that a mixture of 100 μM of IPTG inducing cgRNA B and 100nM OC6 inducing cgRNA D is optimal for switch behavior.

To characterize the toggling behavior, we performed an experiment in four intervals. In the first interval, triggers are expressed to set the switch, setting the fluorescence to high (expressing trigger B) or low (expressing trigger D). In the second interval,

trigger expression is removed to check if the switch remains set, we expect the low fluorescence conditions in interval 1 to remain low fluorescence, and the high fluorescence conditions in interval 1 to remain high fluorescence. In the third interval, trigger expression is inversed to reset the switch in the opposite direction, such that the well with trigger B induction in interval 1 is now induced with trigger D. In the fourth interval, trigger expression is once again removed to check if the reset switch remains in its previous levels. In each interval, cells are grown for 12 hours. At the end of the interval, cells are diluted into fresh media with the proper trigger inducer conditions. In experimental strains, cgRNAs are always induced at the levels previously determined. The high and low fluorescence controls are wells highly induced with cgRNA D (1mM IPTG) and cgRNA B (1μM OC6), respectively and no triggers.

We observe both in time-course plate reader and end-point flow cytometry data that the toggle switch behaves as expected (Figure 2.12). Specifically, the wells initially induced with trigger D have low fluorescence (initially off) and the wells initially induced with trigger B have high fluorescence (initially on) in interval 1. In interval 2, the initially on and initially off wells maintain their high and low fluorescence, respectively. In interval 3, the initially on condition was induced with trigger D and the initially off condition was induced with trigger B such that the initially on condition now has low fluorescence and the initially off condition now has high fluorescence. In interval 4, the wells maintain their low and high fluorescence from the previous interval.

2.4 Discussion

The appeal of using nucleic acids for synthetic genetic networks is two-fold: first, nucleic acids can be programmed to interact in complex and dynamic ways to produce a variety of functions through hybridization; second, multiples of the same nucleic acid interaction pathways can be designed to not interfere with one another, making the scaling of modules possible. These advantages have been thoroughly explored *in vitro* [62], however it has always been difficult to create mechanisms in living cells. The Pierce lab have in the past tried to create a conditional RNAi mechanism, however, there are significant differences between living cells, in lysate, and *in vitro* [11, 12]. In recent years, several nucleic acid regulators have been invented for use in synthetic biology. One of these regulators is the riboswitch [14]. The riboswitch uses a nucleic acid hairpin loop and trigger to activate or inactive translation on an mRNA. While it allows conditional activation or deactivation of

translation it still requires proteins as a circuit intermediary, thus the output of the circuit cannot be orthogonally used to trigger the next stages of the circuit. Another recent nucleic acid regulator is STARs [13]. STARs uses a terminator transcribed prior to the RNA of interest, which can be disturbed by a small RNA trigger, to conditionally activate or inactivate expression. So far, STARs only work in bacteria and do not provide a clean sequestration motif as the trigger binding must occur during transcription. cgRNAs overcome some of these previous limitations in that they work with bacterial and mammalian cells, do not require translation, and have a defined hybridization step after the transcription of both cgRNA and trigger. Furthermore, the triggers for cgRNAs could, in the future, can be appended to or be a part of an existing mRNA, further deepening the integration between synthetic and natural biology.

In this work, we demonstrated the useful features of cgRNA by demonstrating a feedback circuit, with a hybridization based feedback mechanism, and scaling up of that circuit. Previous CRISPR circuits have used miRNA as a means to degrade a RNA transcript to provide feedback [73]. However, the feedback provided in our circuit only made use of hybridization of nucleic acids. We have also not found any demonstrations of the scale up of a circuit schematic utilising the same regulatory components.

In the long run, we hope to achieve truly scalable synthetic genetic circuits, and to move the focus from new modules towards connecting these modules to form ever larger interaction networks. Large gRNA cascades have been demonstrated in the past [72]. However, these cascades do not contain feedback, and they only showed one circuit motif operating in a cell at once. In terms of scalability, the significance of our work is to show two independent circuits operating in the same cell, and to show that we can link these exact same modules together into larger circuits.

We are continually expanding the function and utility of cgRNAs by exploring other Cas variants, such as activator Cas9; building larger libraries of cgRNA regulators; validating cgRNAs with other biological models, such as mammalian cells or embryos; as well as building and validating other circuit motifs that are compatible with cgRNAs. cgRNAs can potentially become a versatile tool for engineering large synthetic genetic networks in the future.

Chapter 3

HCR LATERAL FLOW ASSAYS FOR AMPLIFIED SARS-COV-2 TESTING

Samuel J. Schulte, Jining Huang, and Niles A. Pierce. HCR lateral flow assays for amplified instrument-free at-home SARS-CoV-2 testing, September 2022. URL `https://www.biorxiv.org/content/10.1101/2022.09.18.508442v1`. J.H. participated in the conception of the project, designed, manufactured, and optimized the RNA detection tests, prepared publication data for RNA tests, and participated in preparation of the manuscript.

In March of 2020, the COVID-19 pandemic revealed that lab-based testing could address the technical requirements for detecting the SARS-CoV-2 virus, but could not readily scale to meet the needs of the global population during a pandemic. These initial tests use RT-qPCR. RT-qPCR first reverse transcribes the viral RNA to create a complementary DNA strand, then amplifies that DNA with PCR. During the PCR process, probes attach to newly synthesized DNA and fluoresce, providing signal. RT-qPCR tests can be setup relatively quickly after the publication of the viral genome, however, they require specially stored reagents and laboratory equipment. The lack of scale of RT-qPCR tests caused supply shortages of reagents and equipment all throughout the world causing the need for labs to improvise [82].

In May of 2020, an EUA (emergency use authorization) was given to the first antigen test [83]. However, this test was an immuno-fluorescence test and required specialized equipment to use. The first simple at-home rapid antigen test that reached the scale necessary for the pandemic was the Abbott BinaxNow rapid antigen test released in August of 2020 [84]. The market has since then been flooded by rapid antigen tests; these simple disposable tests could be used at home without special expertise. Rapid anti-gen tests work similarly to pregnancy tests. They employ a lateral flow assay format in which the hCG protein (in the case of a pregnancy test) or the viral nucleocapsid protein (in the case of a rapid SARS-CoV-2 antigen test) moves via capillary forces through a porous substrate, most commonly nitrocellulose, binding in a sandwich between a first antibody carrying a colored label and a second immobilized antibody at the test region to concentrate the analyte–label complex such that it is visible to the naked eye if the target is

sufficiently abundant [107]. The simplicity of these tests makes them ideal for use cases where access to equipment and expertise is limited, such as at-home use.

The large time disparity between the advent of the RT-qPCR test and the rapid antigen tests comes from the ingredients for each test. In the case of the RT-qPCR test, the ingredient specific to the detection of SARS-CoV-2 are the DNA probes used to prime the genomic RNA, the cDNA, and amplified sequences. Once the genomic sequence of the virus is available, DNA probes can be designed and synthesized quickly. Manufacturing of any custom DNA probes is mostly an automated, and fully chemical process that does not require the involvement of any living organisms (other than the human operating the machines). To create a rapid test, however, requires specific antibody pairs that have high affinity to the analyte of interest. To acquire such an antibody pair, the analyte of interest is most often injected into an animal, and the naturally produced antibodies are isolated and screened. In a time of emergent disease, this process of antibody discovery will delay rapid at-home testing of the novel pathogen at a critical time.

Another key issue with rapid antigen tests for SARS-CoV-2 is their relatively high limit of detection. In the first month of pregnancies, hCG concentrations in urine rises to $\approx 10^{10}$ copies/μL [108, 109], and commercial lateral flow assays for pregnancy typically have a limit of detection of $\approx 10^7$ copies/μL. For SARS-CoV-2 tests, the commercial assays detect the N protein. There is a large number of N proteins per virion (≈ 1000) [110]. However, there is a much more limited viral load of virion in a SARS-CoV-2 infection than hCG proteins in pregnancy. In March 2020, two studies found a median viral load of 158 and 3300 virions/μL in saliva [111, 112]. These numbers demonstrate that the lateral flow test format is just barely sensitive enough to detect a SARS-CoV-2 infection. This high limit of detection leads commercial rapid antigen tests to have high false negative rate (e.g., 25–50% in two hospital studies) [85, 86]. Furthermore, in order to shorten the time of assay development previously mentioned, the binding probes must be for a nucleic acid target, since these probes can be designed and manufactured quickly. There is only one copy of the viral RNA genome per virion. In comparison, the RT-qPCR tests can achieve limits of detection of 0.1–0.6 copies/μL of viral RNA genome [113, 114].

In academic circles, there are alternatives to the RT-qPCR and rapid antigen tests. One such type of test uses RT-LAMP [87, 88]. RT-LAMP is similar to a RT-qPCR test, however all nucleic acid synthesis steps are done in the same pot at the same temperature. RT-LAMP accomplishes this by using isothermal enzymes as well

as special primers to form self-priming loops at the end of amplified DNA [89]. The presence of dsDNA changes the pH in the reaction solution, and a sensitive colorimetric pH indicator is used for readout. RT-LAMP improves on the RT-qPCR test by requiring less sophisticated equipment, and less technical lab work while closely maintaining sensitivity. However, it still is used as a laboratory test because it requires enzymes with strict storage conditions, and is subject to the operating requirements of the underlying enzymes.

Another academic alternative is by using CRISPR based readouts for amplified nucleic acids via IVT (*in vitro* transcription). These technologies typically convert RNA or DNA genomes into dsDNA by RPA or RT-RPA, and perform IVT on the dsDNA to synthesize RNA. The RNA is then introduced to a Cas-sgRNA complex, where the sgRNA targets a specific site on the IVT RNA. Once the Cas-sgRNA complex recognizes a site on the IVT RNA, it will collaterally cleave single-stranded nucleic acid sensors with two haptens on each end of the strand. The cleavage or lack of cleavage can be read out via fluorescence or lateral flow. These CRISPR assays once again require less sophisticated equipment, is marginally worse than RT-qPCR in terms of sensitivity, but still require many enzyme steps. While a device [91] that automates this assay is possible, there are still a myriad of different conditions for each step of the test.

We set the goal of developing an amplified lateral flow assay that would enable detection of 1000 SARS-CoV-2 virions/μL, representing an increase in sensitivity of approximately four orders of magnitude relative to at-home pregnancy tests.

To close the gap between our target of 1000 copies of RNA genome per μL and the pregnancy test limit of detection of $\approx 10^7$ copies/μL, we require a large factor of amplification. We hypothesized that signal amplification based on the mechanism of hybridization chain reaction (HCR) [15] would be well-suited for adaptation to the lateral flow assay format.

HCR has previously been used to provide *in situ* signal amplification for multiplexed, spatial imaging for RNA and proteins in fixed biological specimens [16, 17, 20, 21]. The target molecules in a fixed environment are detected by probes carrying HCR initiators and signal is amplified by the polymerization of fluorophore-labelled HCR hairpins, enabling the spatial imaging of molecules within cells, tissue sections, or whole-mount embryos [16–21].

HCR imaging is a multistage process. In the detection stage, probes are added to a fixed sample, and incubated to allow binding. These probes can be DNA or antibodies-conjugated with a DNA initiator. The unused probes are washed in an additional step to reduce non-specific probe binding. In the amplification stage, fluorophore-labelled hairpins are added to the probe-bound sample, and the initiator on the target-bound probes will open the first hairpin, allowing it to expose a second initiator, that in turn opens a second hairpin, that exposes the initiator domain on the probe, enabling polymerization to continue. After polymerization, hairpins are removed via washing to remove non-specific hairpin binding [20]. In the context of fixed samples, a researcher can manually add and wash away the various reagents in the different stages of HCR amplification. However, in order to create a simpler device with as little human intervention as possible, we need to automate reagent delivery and removal, and replace the fluorescence detection scheme with an indicator visible to the human eye.

To automate reagent addition and removal, we planned to use multi-channel lateral flow device designs [115, 116] to sequentially deliver one reagent well at a time in a specified order. To accommodate the lateral flow format, with its comparatively short run time and strict liquid-flow patterns, we worked at higher reagent concentrations than are typical for HCR *in situ* imaging experiments. To eliminate the need for fluorescence imaging equipment, we instead employ hairpins with a conjugated hapten instead of a fluorophore, these haptens can bind to an antibody conjugated to a chromophore to be delivered by a third automated reagent addition step. We do not use hairpins conjugated to chromophore directly because the chromophore is bulkier than hairpins and will interfere with polymerization kinetics.

Seeking to maintain the attractive properties of existing lateral flow tests while addressing the more demanding challenge of SARS-CoV-2 and emergent pathogen detection, we set design criteria:

- *Adaptable*: the creation of a new test for an emergent pathogen should be on the timescale of a week rather than months.
- *Simple*: the test should be as simple to use as a pregnancy test, enabling routine at-home use by a non-expert.
- *Inexpensive*: the test device should be disposable and not require at-home instrumentation.
- *Robust*: the test should avoid reagents (e.g., enzymes) that require cold storage.

- *Rapid*: the test should return results in ≈1 hour, and not days (as is typical of RT-qPCR lab tests).

- *Sensitive*: the test should have a limit of detection of 1000 virions/μL or lower.

Test	Probes	Amplification	Duration (min)
Current work	DNA probes	HCR	90
Commercial A	Proprietary antibody	None	15
Commercial B	Proprietary antibody	None	10
Commercial C	Proprietary antibody	None	15
Commercial D	Proprietary antibody	None	15
Commercial E	Proprietary antibody	None	10

Table 3.1: Descriptions of commercial SARS-CoV-2 rapid tests detecting nucleocapsid protein (N) and amplified HCR lateral flow assay. Description of probes used, amplification techniques and test time for 5 commercial tests and the amplified HCR lateral flow assay described in the current work.

Commercial RT-qPCR offerings are adaptable, inexpensive, and sensitive. Commercial rapid antigen tests (Table 3.1) are simple, inexpensive, robust and rapid. Both the RT-LAMP and CRISPR approaches are potentially adaptable and sensitive. While commercial rapid antigen offerings are highly reliable when they return a positive result (e.g., 96–100%) [85, 86], their lack of sensitivity can lead to a high false negative rate (e.g., 25–50% in two hospital studies) [85, 86].

Test	Gamma-irradiated virus copies/μL							
	100	200	500	1,000	2,000	5,000	10,000	20,000
Current work	✗✗✗	✓✓✓	✓✓✓	✓✓✓	n.t.	n.t.	n.t.	n.t.
Commercial A	✗✗✗	✗✗✗	✗✗✗	✗✗✗	✓✓✓	n.t.	n.t.	n.t.
Commercial B	✗✗✗	✗✗✗	✗✗✗	✗✗✗	✓✓✓	n.t.	n.t.	n.t.
Commercial C	✗✗✗	✗✗✗	✓✓✓	✓✓✓	n.t.	n.t.	n.t.	n.t.
Commercial D	n.t.	n.t.	✗✗✗	✗✗✗	✗✗✗	✗✗✗	✗✗✗	✓✓✓
Commercial E	✗✗✗	✗✗✗	✗✗✗	✓✓✓	✓✓✓	n.t.	n.t.	n.t.

Table 3.2: Amplified HCR lateral flow assay vs five commercial unamplified lateral flow assays. Gamma-irradiated virus spiked into first-channel (this work) or manufacturer-provided (commercial tests) buffer. $N = 3$ replicates for each concentration. Each replicate was judged by eye for a positive (✓) or negative (✗) test result. Not tested (n.t.). See Schulte et al. [117] for images and methodology. Final data collection performed by my collaborator Sam Schulte.

With this work we seek to make an amplified rapid antigen test that is adaptable, but also more sensitive than commercial offerings, while trying to retain the positive characteristics of current rapid antigen tests. In this new testing regime, days, not months after a new emergent pathogen appears, those affected can have a sensitive means to continuously monitor their infection status, and take preventative action based on reliably available data. In developing an amplified HCR lateral flow assay detecting the SARS-CoV-2 RNA genome, we have so far found it necessary to include a heat extraction step, increasing the assay complexity and time, however we achieves a limit of detection lower than five commercial SARS-CoV-2 rapid antigen tests that we evaluated (Table 3.2).

3.1 Device description

To detect viral RNA, we target the same SARS-CoV-2 single-stranded RNA genome that is detected by lab-based PCR tests (Figure 3.1). The target RNA is detected by DNA signal probes complementary to different subsequences along the ≈30,000 nt target, avoiding subsequences shared by other coronaviruses (with the exception of SARS-CoV, which has high sequence similarity to SARS-CoV-2, causes severe disease, and is not circulating [118]). To automatically suppress background that could otherwise arise from non-specific probe binding, our DNA signal probes take the form of split-initiator probe pairs with an HCR initiator split between a pair of probes [20]. As a result, any individual probe that binds non-specifically will not trigger HCR, but specific hybridization of a pair of probes to adjacent cognate binding sites along the target RNA will colocalize a full HCR initiator capable of triggering HCR signal amplification. To maximize sensitivity, the signal probe set comprises 198 split-initiator DNA probe pairs.

By hybridizing to the RNA target, the DNA signal probes create a DNA/RNA duplex at each probe binding site. The probe-decorated target is captured in the test region by an immobilized anti-DNA/RNA capture antibody [119] that binds to DNA/RNA duplexes. Note that while the capture antibody binds DNA/RNA duplexes independent of sequence, any captured RNA that does not include specifically bound split-initiator DNA signal probe pairs will not trigger HCR, and hence will not contribute to background. After immobilization of the probe-decorated target in the test region, colocalized full HCR initiators trigger the self-assembly of DIG-labeled HCR hairpins into HCR amplification polymers decorated with DIG, which are in turn bound by CB-labeled anti-DIG reporter antibodies. Reagents are automatically delivered to the test region in three successive stages using a 3-channel
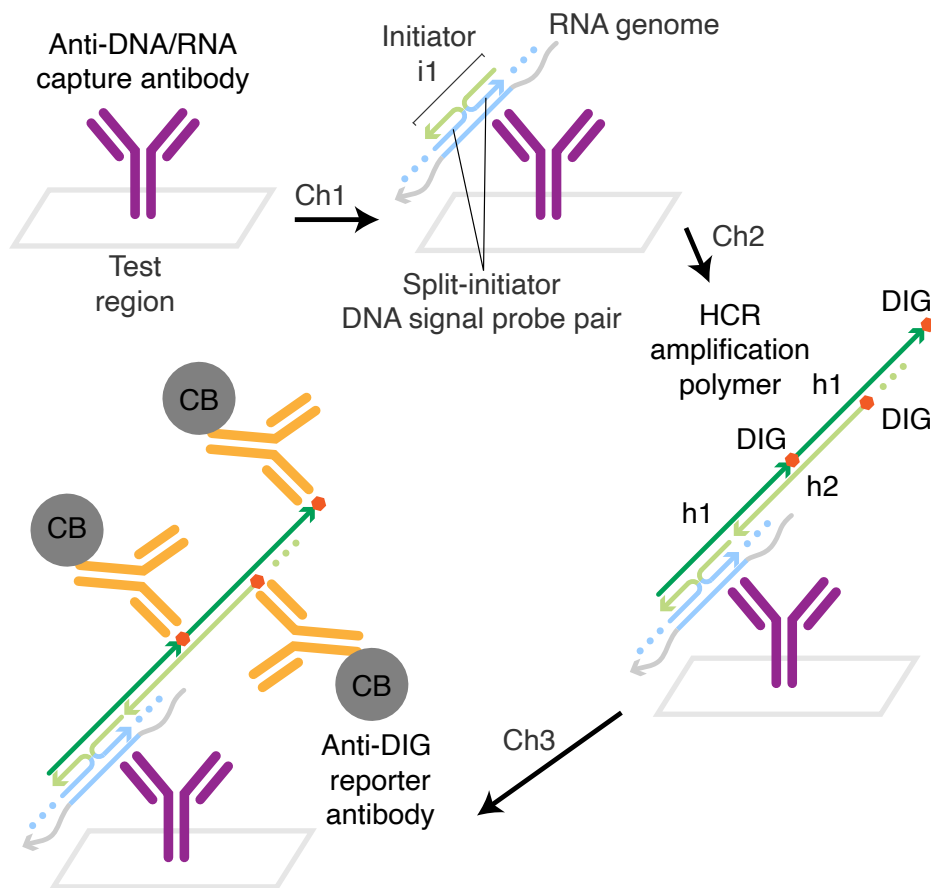
Figure 3.1: Amplified HCR lateral flow assay. Split-initiator DNA signal probe pairs hybridize to cognate binding sites on the viral RNA genome to colocalize full initiator i1; the resulting DNA/RNA duplex is captured in the test region by anti-DNA/RNA capture antibodies; initiator i1 triggers self-assembly of DIG-labeled HCR hairpins (h1 and h2) to form a tethered HCR amplification polymer that is subsequently bound by multiple CB-labeled anti-DIG reporter antibodies, generating multiple units of signal per target. DIG: digoxigenin. CB: carbon black.

membrane (Figure 3.2). The DNA signal probes bind the RNA target and travel via the shortest membrane channel (Channel 1) to reach the test region first, where the probe-decorated target is captured by pre-immobilized anti-DNA/RNA capture antibodies. DIG-labeled HCR hairpins travel via a channel of intermediate length (Channel 2) and reach the test region next, where colocalized full HCR initiators formed by specifically bound split-initiator signal probe pairs trigger growth of tethered DIG-labeled HCR amplification polymers. CB-labeled anti-DIG reporter antibodies travel via the longest channel (Channel 3) and arrive in the test region last,
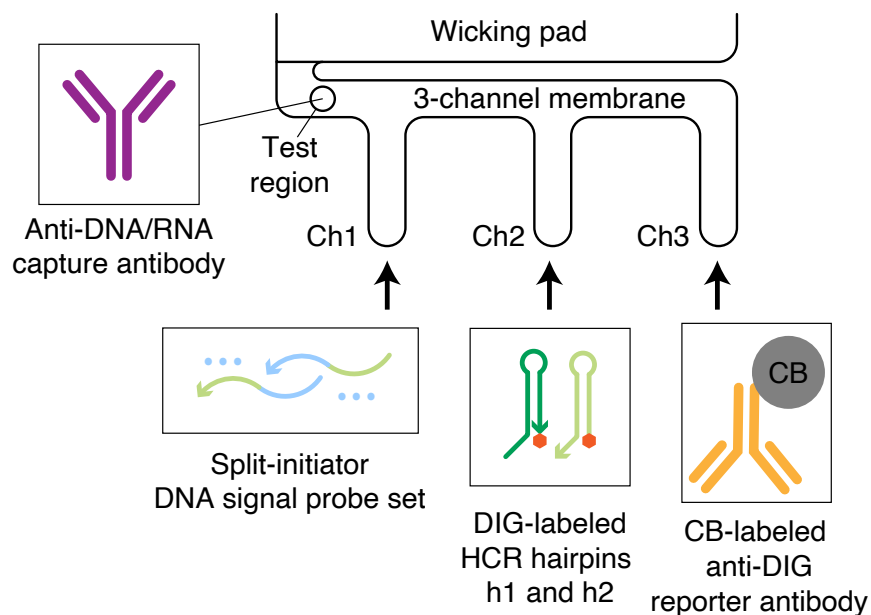
Figure 3.2: Automated delivery of reagents to the test region from Channels 1, 2, and 3 in succession using a 3-channel membrane.

where they bind the HCR amplification polymers to generate an amplified colored signal in the test region.

Extraction of the viral RNA genome requires mild denaturing conditions to remove the viral envelope and the nucleocapsid (N) proteins decorating the RNA genome. These denaturing conditions also help to disrupt native secondary structure in the single-stranded RNA genome prior to detection by the DNA signal probe set. We encountered difficulties incorporating chemical denaturants into the rapid test platform, as the same denaturants that enable target extraction and denaturation also destabilize the DNA/RNA duplexes that underlay target detection. By contrast, heat denaturation can be applied to the sample transiently and then removed to allow DNA probe binding [119]. To date, we have found incorporation of a heat denaturation step to be essential for detection of the SARS-CoV-2 RNA genome in the context of an amplified HCR lateral flow assay. Unfortunately, the addition of a heating step means that we are not yet able to meet our goal that the viral RNA test be as simple to perform as a pregnancy test.

Nonetheless, conceding for the time being that a heat denaturation step is required for the viral RNA lateral flow assay, here we demonstrate a prototype approach using reagents in solution (i.e., using a half-strip assay format) to facilitate heating the sample/probe mixture on a heat block prior to starting the lateral flow assay.

For consistency, the DIG-labeled HCR hairpins and CB-labeled anti-DIG reporter antibodies are also prepared in solution.
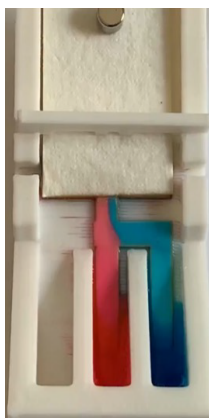
Figure 3.3: Reagents flowing from a forked shape device, reagent from the two separate channels does not mix or flow in sequence, each reagent uses half of the width of the channel after the fork. Experiments performed with food coloring to determine reagent path.

To perform HCR on a nitrocellulose strip, we need to deliver reagents from each of the three sets of reagents consecutively in three separate stages. We used dilute food coloring to visualize liquid flow on various nitrocellulose device configurations. We dipped various devices into wells of food coloring such that flow began in all channels at the same time. We were seeking a device configuration where each reagent will pass through the full width of the common channel, where the test region is, sequentially. In a forked device configuration (Figure 3.3), the flow of both food coloring phases does not mix. Instead, they appear to share the common channel downstream of the fork width wise in two separate phases. While it is desirable for the three stages to not mix prior to the test line, a controlled sequential reagent flow over the entirety of the test line is necessary.

There are examples of a comb configuration of sequential reagent delivery in a lateral flow format in the literature [115, 116]. We designed a three-phase lateral flow device following the comb design pattern (Figure 3.4). The dimensions of our device are influenced by the size of 96-well plates and sufficient separation of the channels to sequentially flow reagents (Figure 3.5). We performed the test with food coloring to ensure reagents would be sequentially delivered and each reagent has full width coverage of the common channel where the test region resides (Figure 3.4).
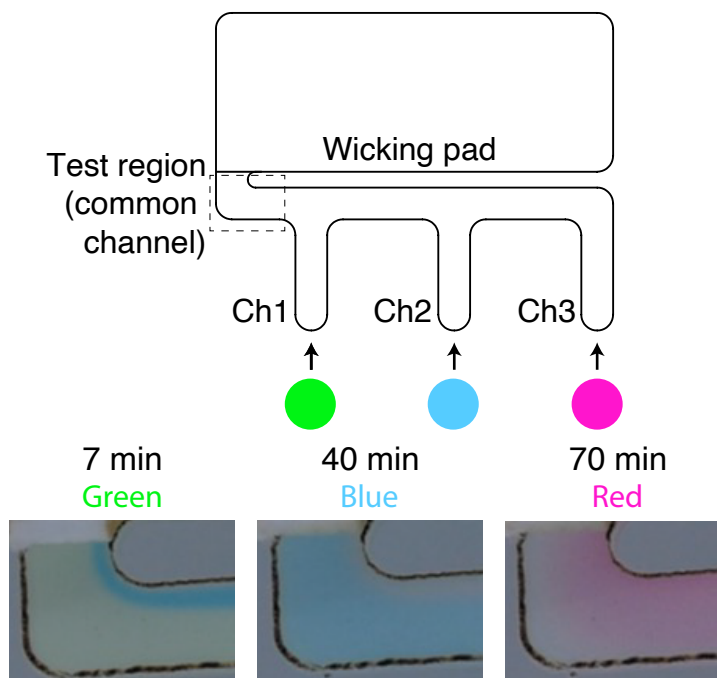
Figure 3.4: Comb shaped device enables sequential reagent steps over test line. Experiments performed with food coloring to determine reagent path. Top: visualization of the comb device to sequentially deliver all reagents across the entire width of the combined channel. Green, blue, and red food coloring is flown into the first, second, and third channels, respectively. Bottom: visualization of sequential flow of reagents through the common channel in a comb format using food coloring.

To perform a test, the sample is added to extraction buffer containing the DNA signal probes (Channel 1 reagents), heated to 65 °C for 15 minutes, and then loaded into a well (Channel 1 well) on a 96-well plate proximal to wells containing the DIG-labeled HCR hairpins (Channel 2 well) and CB-labeled anti-DIG reporter antibodies (Channel 3 well; Figure 3.2). To start the lateral flow assay, the ends of the three membrane channels are simultaneously submerged into the three wells (Figure 3.6), leading to automated successive delivery of the Channel 1, 2, and 3 reagents to the test region without user interaction. After 90 minutes, the result is read as either a positive result (black signal) or a negative result (no signal) in the test region with the naked eye (Figure 3.7).

The 90-minute duration of this lateral flow assay is because the spacing between the wells on a 96-well plate dictated long channels, and also because our use of larger reagent volumes delayed the transitions between channel flows (Figure 3.4).
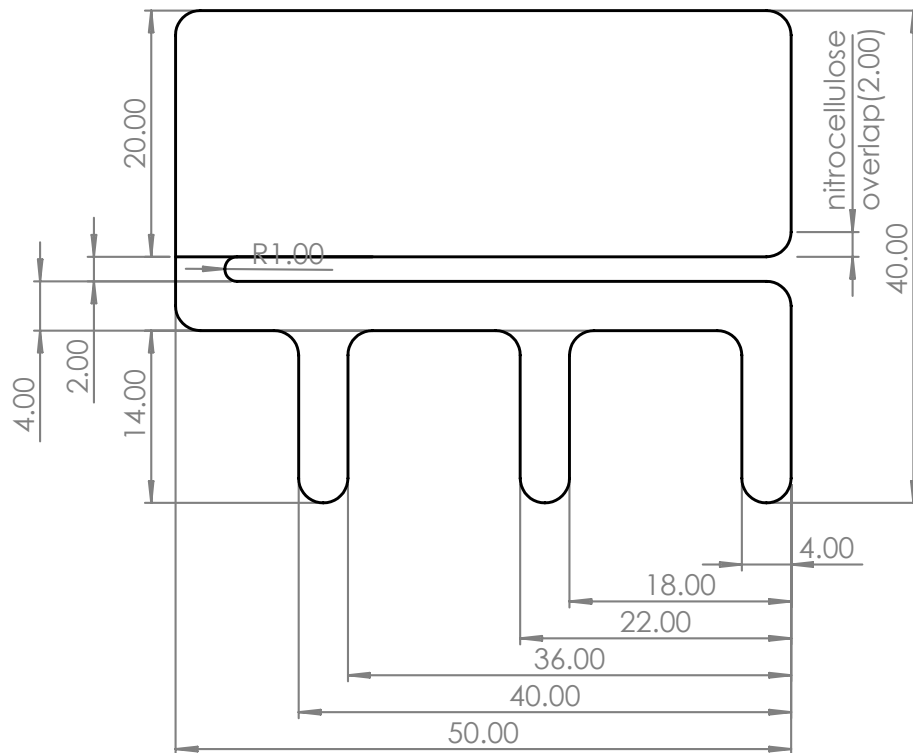
Figure 3.5: Nitrocellulose membrane and wicking pad dimensions for viral RNA detection. The nitrocellulose membrane and wicking pad were overlapped on a polyethylene backing card and cut to size with a laser cutter. All dimensions are shown in millimeters (R: radius). All curved lines have a radius of 2.00 mm unless otherwise indicated.

Figure 3.6: Photograph of the lateral flow device for viral RNA detection. The lateral flow device with three channels resting within their respective wells on a 96-well plate.

Gamma-irradiated SARS-CoV-2 (copies/µL)

| 100 | 200 | 500 | 1000 |
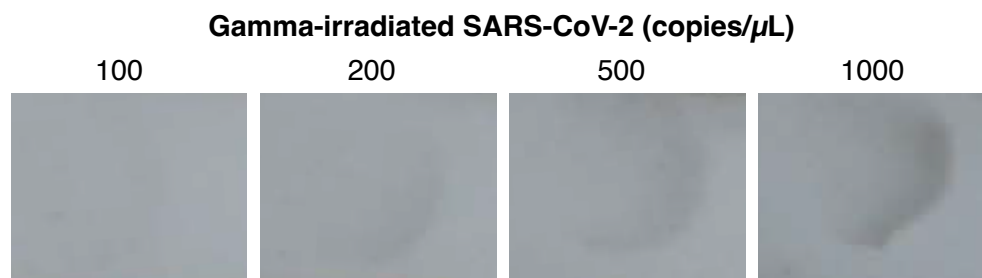
Figure 3.7: Sensitivity: gamma-irradiated SARS-CoV-2 spiked into extraction buffer at different concentrations revealed a limit of detection of 200 copies/µL.

Off-target viral RNA genomes

| No virus | 229E | HKU1 |

Figure 3.8: Cross-reactivity: no staining observed for extraction buffer spiked with synthetic RNA genomes from different coronaviruses 229E (7,200 copies/µL) or HKU1 (10,000 copies/µL).

3.2 Test characterization

To characterize sensitivity, gamma-irradiated SARS-CoV-2 virus was spiked into extraction buffer with DNA signal probes, heated to 65 °C for 15 min, and run on the amplified HCR lateral flow assay, revealing a limit of detection of 200 virions/µL (Figure 3.7). No background staining was observed in the test region for experiments run without spiked-in virus (Figure 3.8). To characterize cross-reactivity, experiments were run with synthetic RNA genomes from other coronaviruses spiked in at high concentration (7,200 copies/µL for 229E and 10,000 copies/µL for HKU1); no staining was observed in the test region in either case (Figure 3.8). By achieving a limit of detection of 200 copies/µL, this viral RNA test is more sensitive than all five commercial lateral flow assays that we evaluated for viral protein detection, with the caveat that this RNA test uses a half-strip assay format (with reagents in solution rather then dried onto conjugate pads). For the RNA detection setting, the HCR amplification gain was 10 ± 3 (mean ± estimated standard error of the mean via uncertainty propagation for $N = 3$ replicate assays for each experiment type;

Figure 3.9 and Table 3.3), measured by comparing the signal intensity for assays run using both HCR hairpins (h1 and h2) to assays run using only hairpin h1.
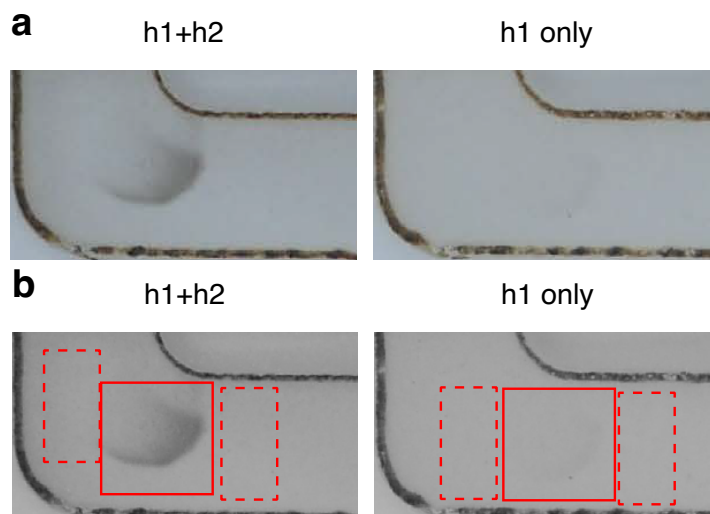
Figure 3.9: Gamma-irradiated virus was spiked into extraction buffer at 5000 copies/μL for all tests. N = 3 replicate assays for unamplified and amplified HCR lateral flow test. (a) Raw images. (b) Images after conversion to gray-scale. Quantitative image analysis following the methods of Schulte et al. [117] using the depicted signal boxes (solid boundary) and background boxes (dashed boundary).

	Signal$_{h1+h2}$	Signal$_{h1}$	Amplification gain
Viral RNA detection assay	13 ± 2	1.3 ± 0.3	10 ± 3

Table 3.3: Estimated amplification gain in the context of lateral flow assays for viral RNA detection. Quantitative image analysis of the amplified (h1 and h2) and unamplified (h1 only) assays of Figure 3.9 following the methods of Schulte et al. [117]. Mean ± estimated standard error of the mean via uncertainty propagation for N = 3 replicate assays for each experiment type.

To demonstrate the usefulness of this lateral flow test platform during an emergent disease, we swapped our SARS-CoV-2 probes for probes targeting the single stranded RNA genome for the H3N2 influenza virus. After we received the probes, it took 4 days of laboratory work to purify the probes via PAGE purification, and performed the new H3N2 assay. We were able to detect synthetic H3N2 RNA genome to 500 copies/μL in the first experiment after acquiring the probes (Figure 3.10).

RNA detection with on-strip HCR amplification introduces a new approach to lateral flow testing for infectious disease, as to our knowledge, all current unamplified

Synthetic H3N2 Genome (copies/µL)

| 200 | 500 | 1000 | 10000 |

Figure 3.10: Sensitivity: synthetic H3N2 genomic RNA spiked into extraction buffer at different concentrations revealed a limit of detection of 500 copies/µL (N=1).

commercial lateral flow assays detect protein rather than nucleic acids. An important benefit of RNA detection is the ability to design and s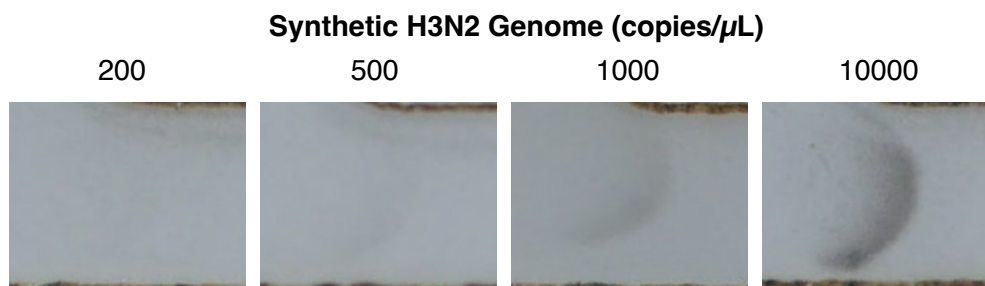ynthesize a new DNA signal probe set within a matter of days following sequencing of a new RNA target of interest. As a consequence, shortly after the emergence of a new pathogen, an amplified HCR lateral flow assay could be developed and tested. This capability would fill a critical gap in the at-home testing infrastructure while antibodies suitable for a protein-detection lateral flow assay are developed and screened.

Despite not fully fulfilling the simplicity criteria, viral RNA detection with an amplified HCR lateral flow assay demonstrates promise. Currently, the addition of a heating step undermines our goals in several regards, reducing simplicity by adding an extra step, requiring the use of a heat block (which can be viewed as a primitive form of instrumentation), and increasing the test duration. Further testing of chemical denaturants, as well as strategies for automated removal of denaturants, may eliminate the need for a heat denaturation step, enhancing the assay in multiple regards. Nonetheless, with the exception of the heating step, the assay remains simple, exploiting a 3-channel membrane to automatically perform HCR signal amplification without user interaction. Sensitivity is enhanced compared to unamplified commercial SARS-CoV-2 viral protein tests, achieving a limit of detection for gamma-irradiated SARS-CoV-2 of 200 copies/µL. The reagents (anti-DNA/RNA capture antibody, DNA signal probes, DIG-labeled HCR hairpins, CB-labeled anti-DIG reporter antibodies) are inexpensive, and only the human eye is needed to read out the result. If the heating step can be removed, it appears feasible to meet all six design requirements using a full-strip lateral flow assay format (with reagents dried onto conjugate pads). Even if the heating step cannot be removed, RNA-detection HCR lateral flow assays may still prove useful due to their suitability for rapid deployment of new tests in the face of emerging pathogens.

3.3 Discussion

At the outbreak of a pandemic, an assay that can continuously test and monitor the population could be transformative in preventing infectious disease transmission. For example, data and modeling suggest that more than half of SARS-CoV-2 infections are spread unknowingly by asymptomatic carriers [120, 121]. To enable this continuous testing, we developed an amplified HCR lateral flow assay for viral RNA detection that is inexpensive (readout via the naked eye), robust (enzyme-free), adaptable, and sensitive (detecting 200 copies/μL of gamma-irradiated SARS-CoV-2). By comparison, five unamplified commercial lateral flow assays targeting the N protein have limits of detection that are 2.5$\times$ to 100$\times$ higher than our amplified HCR assay, not to mention the prolonged 7-month period where no rapid at-home tests were available. Commercial unamplified SARS-CoV-2 rapid tests lead to high false-negative rates [85, 86], thus any gain in sensitivity is in a critical regime where a difference can be made in terms of false negative test results, and in turn the number of active disease carriers at the onset of an emergent infectious disease.

We also demonstrate it is possible to combine the signal amplification of a HCR-based assay with the simplicity of a lateral flow assay. This was achieved using a 3-channel membrane to sequentially deliver reagents with no user interaction after initiating the test. Further optimizations can be made to this test in terms signal strength, simplicity, and versatility of the assay. There is potential to increase the HCR signal gain from one order of magnitude to the two orders seen in HCR imaging applications [21]. The heating step prevents this current device from being as simple as a pregnancy or a commercial rapid antigen test. Further optimization should be made to either chemically or physically remove this step or automate this step.

While the run time of the amplified test described is higher than most commercial rapid antigen tests, we anticipate that in many situations a more sensitive test will be preferred. By switching out SARS-CoV-2 probes for probes targeting other pathogens, amplified HCR lateral flow assays offer a versatile platform for sensitive testing, including emerging pathogens. Amplified HCR lateral flow assays for viral RNA detection have not yet achieved our goals for simplicity, but represent a promising new approach for infectious disease testing and would enable nimble assay development upon the sequencing of novel pathogens.

Chapter 4

NUPACK: ANALYSIS AND DESIGN OF NUCLEIC ACID STRUCTURES, DEVICES, AND SYSTEMS

We are engaged in a multi-decade effort to develop NUPACK (Nucleic Acid Package), a growing software suite for the analysis and design of nucleic acid structures, devices, and systems [122]. NUPACK algorithms [5, 7, 8, 92, 123–126] are formulated in terms of nucleic acid secondary structure (i.e., the base-pairs of a set of DNA or RNA strands) and employ empirical free energy parameters [2, 3, 127–137].

Problem categories

NUPACK algorithms address two fundamental classes of problems:

- *Sequence analysis:* given a set of DNA or RNA strands, analyze the equilibrium base-pairing properties over a specified ensemble.
- *Sequence design:* given a set of desired equilibrium base-pairing properties, design the sequences of a set of DNA or RNA strands over a specified ensemble. Sequence design is performed subject to diverse sequence constraints.

Ensembles

NUPACK algorithms operate over two fundamental ensembles:

- *Complex ensemble:* the ensemble of all (unpseudoknotted connected) secondary structures for an arbitrary number of interacting RNA or DNA strands.
- *Test tube ensemble:* the ensemble of a dilute solution containing an arbitrary number of RNA or DNA strand species (introduced at user-specified concentrations) interacting to form an arbitrary number of complex species.

Furthermore, to enable reaction pathway engineering of dynamic hybridization cascades (e.g., shape and sequence transduction using small conditional RNAs [1, 11]) and large-scale structural engineering including pseudoknots (e.g., RNA origamis [138]), NUPACK generalizes sequence analysis and design to *multi-tube ensembles* comprising one or more test tubes [8].

Note that a complex ensemble is subsidiary to a test tube ensemble, so complex analysis is inherent in test tube analysis (but not vice versa), and complex design is inherent in test tube design (but not vice versa). As it is typically infeasible to experimentally study a single complex in isolation, we recommend analyzing and designing nucleic acid strands in a test tube ensemble that contains the complex of interest as well as other competing complexes that might form in solution. For example, if one is experimentally studying strands A and B that are intended to predominantly form a secondary structure within the ensemble of complex A·B, one should not presuppose that the strands do indeed form A·B and simply analyze or design the base-pairing properties of that complex. Instead, it is more physically relevant to analyze or design a test tube ensemble containing strands A and B interacting to form multiple complex species (e.g., A, B, A·A, A·B, B·B) so as to capture both concentration information (how much A·B forms?) and structural information (what are the base-pairing properties of A·B when it does form?).

All-new cloud-based NUPACK web app

Since its launch in 2007, usage of the NUPACK web app (nupack.org) [122] has increased to the point where the underlying static compute cluster is frequently overwhelmed by user demand. To provide a scalable resource for the global research community, the NUPACK web app was re-architected from the ground up to exploit a scalable compute cluster that resizes dynamically in the cloud in response to user demand. The all-new NUPACK web app integrates diverse components to create an intuitive and powerful analysis and design environment:

- *Algorithms:* mathematically rigorous, physically sound, computationally efficient scientific algorithms [5, 7, 8, 92, 123–126].
- *Hardware:* a hybrid cloud compute cluster combining local hardware and scalable cloud hardware.
- *Interface:* an intuitive web interface for rapid job submission and result inspection.

- *Graphics:* publication-quality client-side graphics to enable straightforward interpretation of results, interactive data, and efficient preparation of talks and papers.

Researchers can run jobs and inspect results within the NUPACK web app, which leverages NUPACK 4 scientific code base as its backend.

Outline

We summarize the features of the Analysis, Design, and Utilities pages of the NUPACK web app:

- *Analysis page:* Analyze the equilibrium base-pairing properties of one or more test tube ensembles (and subsidiary complex ensembles). These are the all-purpose sequence analysis tools.
- *Design page:* Design the sequences for one or more test tube ensembles (and subsidiary complex ensembles). These are the all-purpose sequence design tools.
- *Utilities page:* Analyze, design, or prepare figures for a single complex ensemble. These are quick tools applicable when your ensemble is a single complex.

We further describe the control plane that enables the all-new NUPACK web app to burst dynamically in the cloud. We then characterize the performance of the old vs. new NUPACK web app and the performance uplifts from bursting into the cloud.

4.1 Analysis page

The Analysis page of the NUPACK web app (Figure 4.1) enables users to analyze the equilibrium concentration and base-pairing properties of a multi-tube ensemble containing one or more test tubes. Each test tube ensemble contains a user-specified set of strand species, each introduced at a user-specified concentration.

Input

The Analysis Input page allows the user to specify the physical model and components for the multi-tube ensemble. For the multi-tube ensemble, specify the following:

- *Material:* Select RNA or DNA.

Figure 4.1: NUPACK analysis page.

- *Temperature:* Specify the temperature in Celsius (or select "Melt" and specify a minimum temperature, increment, and maximum temperature to simulate the multi-tube ensemble for a range of temperatures).

- *Model options:* Optionally specify details of the physical model:

  - *Parameters:* Select from available free energy parameter sets.

  - *Ensemble:* Specify the coaxial and dangle stacking formulation.

  - *Salts:* Specify salt concentrations (Na^+ and Mg^{++}).

For each test tube within the multi-tube ensemble, specify the following:

- *Tube name:* Specify the name of the test tube.

- *Strands:* Specify the name, sequence, and concentration of each strand species.

- *Complexes:* Specify the complex species in the test tube in any of three ways:

  - *Max complex size:* Automatically generate all complexes up to a specified maximum number of strands (default: 1).

  - *Include complex:* explicitly specify complexes to include in the test tube ensemble (that would otherwise not be included based on the specified maximum complex size).

  - *Exclude complex:* explicitly specify complexes to exclude from the test tube ensemble (that would otherwise be included based on the specified maximum complex size).

Computation

For each complex in the multi-tube ensemble, the partition function, equilibrium base-pairing probabilities, and minimum free energy (MFE) proxy structure are calculated using a unified dynamic programming framework [5, 92]. If the same strand species are present in more than one tube of a multi-tube ensemble, the algorithms achieve significant cost savings relative to analyzing each tube in the ensemble separately [92]. For each test tube ensemble, the equilibrium complex concentrations are the solutions to a strictly convex optimization problem (formulated in terms of calculated partition functions and user-specified strand concentrations), which we solve efficiently in the dual form [5]. The equilibrium complex concentrations and base-pairing probabilities are then used to calculate the test tube fraction of bases unpaired and test tube ensemble pair fractions [5]. In order to analyze the tube properties for a different set of strand concentrations, it is not necessary to rerun the dynamic programs to calculate partition functions, equilibrium base-pairing probabilities, and MFE proxy structures; only the convex optimization problem must be solved again, and this can be rapidly done from within the Results page.

Results

The Analysis Results page summarizes the equilibrium concentration and base-pairing properties of each test tube in the multi-tube ensemble; use the dropdown to examine the results for a given tube.

- *Temperature slider:* For calculations that specified a temperature range, use the temperature slider to examine results over the range of simulated temperatures.

- *Melt profile:* For calculations that specified a temperature range, the melt profile depicts the equilibrium fraction of bases unpaired in the test tube as a function of temperature.

- *Equilibrium complex concentrations:* The bar graph depicts the equilibrium concentration of each complex that forms with appreciable concentration in the test tube (adjust the display filters to alter which complex concentrations are shown). Clicking any bar to display equilibrium base-pairing information for the corresponding complex:

  - *MFE structure:* Depicts the MFE proxy structure for the complex. In the default view, each base is shaded with the probability that it adopts the depicted base-pairing state at equilibrium, revealing which portions of the structure usefully summarize equilibrium structural features of the complex ensemble.

  - *Pair probabilities:* Depicts equilibrium base-pairing probabilities for the complex. By definition, these data are independent of concentration and of all other complexes in solution. The area and color of each dot scale with the equilibrium probability of the corresponding base pair. With this convention, the matrix is symmetric, as denoted by a diagonal line. In the column at right, the area and color of each dot scale with the equilibrium probability that the corresponding base is unpaired within the complex ensemble. Optional black circles depict each base pair or unpaired base in the MFE proxy structure.

- *Test tube ensemble pair fractions:* Depicts equilibrium base-pairing information for the test tube ensemble, taking into account the equilibrium concentration and base-pairing properties of each complex. The area and color of the dot at row i and column j scale with the equilibrium fraction of base i

that is paired to base j in solution. With this convention, the matrix can be asymmetric. In the column at right, the area and color of the dot in row i scales with the equilibrium fraction of base i that is unpaired within the test tube ensemble.

Information can be exported to the Design or Utilities pages:

- *To Design:* For a given complex, export the MFE proxy structure to the Design page to redesign the sequence.

- *To Utilities:* For a given complex, export the MFE proxy structure and sequence information to the Utilities page to annotate publication quality graphics, or to do quick analysis or design calculations in the context of the complex ensemble.

For individual plots, download graphics for editing in vector graphics programs or download data for local plotting. Alternatively, all job data and plots can be downloaded as a single compressed file.
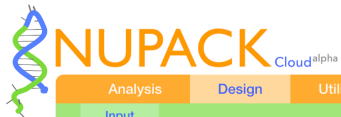
4.2   Design page

The Design page of the NUPACK web app (Figure 4.2) allows users to perform sequence design over a multi-tube ensemble comprising one or more target test tubes. Wolfe et al. [8] describe the design formulation and details on how to formulate a multi-tube design problem.

Input

The Design Input page allows the user to specify the physical model and components for the multi-tube ensemble. For the multi-tube ensemble, specify the following:

- *Material:* Select RNA or DNA.

- *Temperature:* Specify the temperature in Celsius.

- *Trials:* Specify the number of independent design trials.

- *Model options:* Optionally specify details of the physical model:

  - *Parameters:* Select from available free energy parameter sets.

  - *Ensemble:* Specify the coaxial and dangle stacking formulation.

Figure 4.2: NUPACK design page.

- *Salts:* Specify salt concentrations (Na^+ and Mg^{++}).

- *Algorithm settings:* Optionally specify algorithm settings:

  - *Stop condition:* Specify a stop condition in the range (0,1). The design algorithm will attempt to reduce the augmented objective function (weighted ensemble defect plus weighted soft constraints) below the stop condition while satisfying hard constraints.

  - *Max design time:* Specify the maximum design time.

  - *Random seed:* Specify a non-zero integer for a reproducible design trial (default: 0; corresponding to a random trial).

  - *Wobble mutations:* For RNA designs, globally prohibit (default) or allow wobble mutations (yielding G·U base pairs). Note that for RNA designs, wobble mutations can also be allowed locally when specifying complementarity hard constraints.

Specify all components that appear in one or more target test tubes within the multi-tube ensemble:

- *Domains:* Specify sequence domains. A domain is a set of consecutive nucleotides that appear as a subsequence of one or more strands in the design, specified as a name and a sequence (specified 5′ to 3′ using degenerate nucleotide codes. Note that specification of a domain using degenerate nucleotide codes represents an implicit hard sequence constraint.

- *Strands:* Specify target strands. Each target strand is a single RNA or DNA molecule specified as a name and a sequence (specified 5′ to 3′ in terms of previously specified domains).

- *Target complexes:* Specify target complexes. Each target complex is an on-target and/or off-target complex specified as a name and an ordered list of strands (i.e., an ordering of strands around a circle in a polymer graph) and a complex name. If the complex is to be used as an on-target complex in at least one target test tube, it is specified with an on-target secondary structure; the target structure will be ignored in target test tubes where a complex appears as an off-target complex.

For each target test tube in the multi-tube ensemble, specify the following:

- *Target tube name:* Specify the name of the target test tube.

- *On-target complexes:* Specify a set of desired on-target complexes (from the previously specified set of target complexes that include a target secondary structure), each with a target concentration.

- *Off-target complexes:* Specify undesired off-target complexes in any of three ways:

  - *Max complex size:* Automatically generate the set of all off-target complexes up to a specified maximum number of strands (default: 1).

  - *Include complex:* Explicitly specify off-target complexes to include in the test tube ensemble (that would otherwise not be included based on the specified "Max complex size").

  - *Exclude complex:* Explicitly specify off-target complexes to exclude from the test tube ensemble (that would otherwise be included based the specified "Max complex size").

Note that any complex included as an on-target complex will not be included as an off-target complex. Note also that if an off-target is specified using a target complex for which a target structure has been specified, the target structure is ignored (by definition, there is no target structure for an off-target complex). Note further that used together, "Max complex size" and "Exclude complex" provide a powerful combination for specifying target test tubes. With "Max Complex Size" it is possible to specify a large set of off-target complexes formed from a set of system components. With "Exclude complex" it is further possible to remove from this large set all of the cognate products that should form between these system components (so they appear as neither on-targets nor off-targets in the tube ensemble). For example, with this approach, the reactive species in a global crosstalk tube can be forced to either perform no reaction (remaining as desired on-targets) or to undergo a crosstalk reaction (forming undesired off-targets), enabling minimization of global crosstalk during sequence optimization.

Optionally specify hard constraints, soft constraints, and/or defect weights for the multi-tube ensemble:

- *Hard constraints:* Specify hard constraints that prohibit sequences that violate the constraints, including match constraints, complementarity constraints,

diversity constraints, similarity constraints, window constraints, library constraints, and pattern constraints.

- *Soft constraints:* Specify soft constraints that penalize (but do not prohibit) suboptimal sequences, including similarity constraints, pattern constraints, symmetry constraints, and energy match constraints.

- *Defect weights:* Specify defect weights to prioritize or de-prioritize design quality for any combination of domain, strand, complex, or tube. Note that a defect weight can be specified as either an absolute weight or as a multiplier of existing weights.

Computation

The sequence design algorithm seeks to iteratively reduce the augmented objective function (weighted multi-tube ensemble defect plus weighted soft constraints) below a stop condition while satisfying the specified hard constraints. During sequence optimization, candidate mutations to a random initial sequence are efficiently evaluated over the multi-tube ensemble by estimating the multi-tube ensemble defect using test tube ensemble focusing, hierarchical ensemble decomposition, and conditional physical quantities calculated within subensembles [7, 8, 126]. The progress page displays, for each independent design trial, the augmented objective function as a function of design time.

Results

The following two plots summarize the design results for each independent design trial:

- *Augmented objective function:* This plot displays, for each independent design trial, the augmented objective function comprising:

    - The *weighted objective function*, incorporating any defect weights specified by the user. With the default value of unity for all weights, this reduces to the multi-tube ensemble defect, representing the average equilibrium fraction of incorrectly paired nucleotides over the multi-tube ensemble.

    - The *weighted soft constraint* contribution for each soft constraint type specified by the user.

- *Multi-tube ensemble defect:* This plot displays, for each independent design
trial, the multi-tube ensemble defect, representing the average equilibrium
percentage of incorrectly paired nucleotides over the multi-tube ensemble.
For each design trial, the defect contributions within the multi-tube ensemble
come in two varieties:

  - The *structural defect* component quantifies the fraction of nucleotides
  that are in the incorrect base-pairing state within the correct complex.

  - The *concentration defect* component quantifies the fraction of nucleotides
  that are in an incorrect base-pairing state because they are not in the cor-
  rect complex.

Click on the bar for any design trial to explore details for that design trial:

- *Tube defects:* This plot displays, for each target test tube, the *test tube ensem-
ble defect*, representing the equilibrium concentration of incorrectly paired
nucleotides over the ensemble of the test tube. For each target test tube, the
defect contributions come in two varieties:

  - The *structural defect* component represents the equilibrium concentra-
  tion of nucleotides that are in the incorrect base-pairing state within the
  correct complex.

  - The *concentration defect* component represents the equilibrium concen-
  tration of nucleotides that are in an incorrect base-pairing state because
  they are not in the correct complex.

- *Sequences:* Sequence design results are displayed for each sequence domain
and each strand in the design ensemble.

Click on the bar for any tube to explore details for that tube:

- *On-target complex contribution to tube defect:* This plot displays, for each on-
target complex, the contribution to the test tube ensemble defect, representing
the equilibrium concentration of incorrectly paired nucleotides over the en-
semble of the test tube. For each on-target complex, the defect contributions
come in two varieties:

- The *structural defect* component represents the equilibrium concentration of nucleotides that are in the incorrect base-pairing state within the ensemble of the complex.

- The *concentration defect* component represents the equilibrium concentration of nucleotides that are in an incorrect base-pairing state because there is a deficiency in the concentration of the complex.

- *On-target complex defect:* This plot displays, for each on-target complex in the test tube, the *complex ensemble defect*, representing the equilibrium number of incorrectly paired nucleotides over the ensemble of the complex.

- *On-target complex concentration:* This plot displays, for each on-target complex in the test tube, the *equilibrium complex concentration* and the target concentration.

- *Off-target complex concentration:* This plot displays the equilibrium complex concentration for each off-target complex that forms appreciably in the test tube.

Click on the bar for any on-target complex to explore details for that complex:

- *Target structure:* Depicts the target secondary structure for the on-target complex. By default, each base is shaded with the probability that it adopts the depicted base-pairing state at equilibrium within the complex ensemble. Optionally, each base is shaded according to its identity.

- *Pair probabilities:* Depicts equilibrium base-pairing probabilities for the on-target complex. By definition, these data are independent of concentration and of all other complexes in solution. The area and color of each dot scale with the equilibrium probability of the corresponding base pair. With this convention, the matrix is symmetric, as denoted by a diagonal line. In the column at right, the area and color of each dot scale with the equilibrium probability that the corresponding base is unpaired within the complex ensemble. Optional black circles depict each base pair or unpaired base in the target structure.

Information can be exported to the Analysis or Utilities pages:

- *To Analysis:* For the multi-tube ensemble, a given target test tube, or a given on-target complex, export the designed sequences to the Analysis page for further equilibrium analysis.

- *To Utilities:* For a given on-target complex, export the target structure and designed sequences to the Utilities page to annotate publication quality graphics, or to do quick analysis or design calculations in the context of the complex ensemble.

For individual plots, download graphics for editing in vector graphics programs or download data for local plotting. Alternatively, all job data and plots can be downloaded as a single compressed file.

4.3   Utilities page

The Utilities page of the NUPACK web app (Figure 4.3) allows users to analyze, design, or annotate the equilibrium properties of a complex. The page accepts as input either sequence information, structure information, or both, performing diverse functions based on the information provided, including:

- Evaluation and display of equilibrium base-pairing information for a specified secondary structure in the context of the complex to which it belongs.
- Automatic layout, rendering, and annotation of secondary structures.
- Sequence analysis or design for a complex ensemble.

For individual plots, download graphics for editing in vector graphics programs or download data for local plotting. Alternatively, all job data and plots can be downloaded as a single compressed file.

Information can be exported to the Analysis or Design pages:

- *To Analysis:* For the specified complex, export the strand sequences to the Analysis page to analyze in the context of a test tube ensemble.

- *To Design:* For the specified complex, export the specified structure to the Design page to design the sequence in the context of a test tube ensemble, carrying along any specified sequence constraints.

Note that for a given complex ensemble:

Figure 4.3: NUPACK utilities page.

- The Analysis page displays results through the lens of the MFE proxy structure.

- The Design page displays results through the lens of the target structure.

- The Utilities page displays results through the lens of a user-specified structure.

4.4 Control plane and infrastructure

From a user friendliness perspective, the new NUPACK web application will use a cluster burstable to the cloud to alleviate over-utilization issues and allow users to keep a history of their previous jobs. The previous NUPACK implementation used an off-the-shelf scheduler, and ran web code intermingled with drawing and dispatch code on the same head node. This approach is not scalable, the scheduler cannot incorporate new cloud nodes when the cluster needs to burst, and the head node can become overwhelmed. Furthermore, the technical debt and the difficulty of maintenance has reached a critical level where the web app needs to be rearchitected from the ground up. In our new approach, all components of the application can be ran in a distributed way, enabling horizontal scaling of the control plane as well as the workers. All components of the application are also containerized, allowing future maintainers to easily deploy and reassemble all aspects of NUPACK. The new implementation of NUPACK consists of three major components, the frontend, control plane, and the workers. Both the frontend and the control plane will be hosted on AWS (Amazon Web Services) to ensure robustness and reliability.

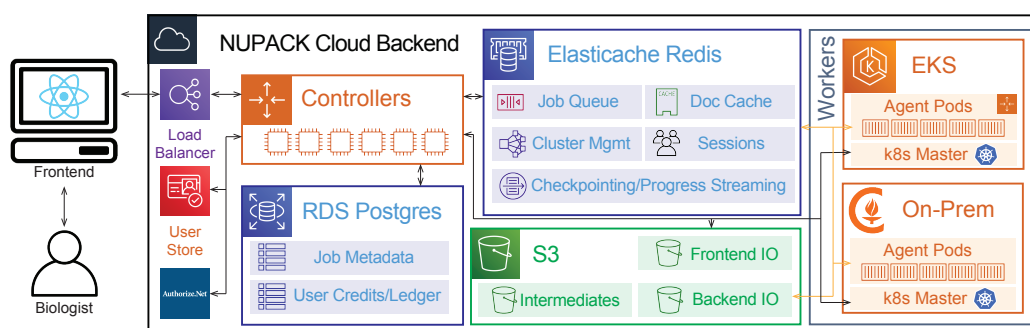
Figure 4.4: NUPACK backend architecture.

Placement and execution of jobs

The control plane for the new NUPACK web application can be divided into four components (Figure 4.4): 1) the controllers, responsible for coordination of jobs and worker containers, 2) ElastiCache Redis, responsible for caching and communications, 3) S3, the job storage system, and 4) RDS Postgres, the job metadata

system. Each component will be described in detail in this section. The workers will be described in the next section.

- **Controllers:** the controllers coordinate all jobs as well as the placement of worker containers. The controllers can be distributed and do not require a leader in its cohort. Controllers split one of three frontend job types into multiples of three sub-job types. The three frontend job types are: analysis, design, and utilities, where utilities can be either a complex analysis or design job. The sub-jobs calculate: pair probabilities (pairs sub-jobs), MFE structure (MFE sub-jobs) or a single design trial (design sub-jobs). An analysis job with N tubes is typically split into N pairs sub-jobs and N MFE sub-jobs, each with a list of complexes to be analyzed. A design job is typically split into individual trials, each trial is reseeded, and each trial is packaged as a design sub-job. A utilities job can be split into a combination of pairs, MFE, or design sub-jobs based on the user's input and calculations selected. Each sub-job is pushed to the job queue and performed by the workers. The controllers proxy any intermediate results or progress updates back to the user. Once all sub-jobs are complete, the controllers retrieve the frontend job specifications, the sub-job results and compiles frontend job results that are sent to the user. Each of these steps touches on the various parts of the backend described below.

- **ElastiCache Redis:** The communications and caching for NUPACK are performed by Redis. Sub-jobs are sent to the queue as a ranked list with their insertion time as the rank. They are processed in order by the worker containers. Once a worker container begins a job, it sends progress updates to Redis via a stream. There are three types of signals a worker node can send as progress updates: a heartbeat signal to indicate the job is currently running or finished, a batched signal to indicate the number of complexes complete compared to the total number of complexes in the job, and a design defect signal to indicate the current design defect. These progress signals are used by the API to trigger further job actions and are sent to the client to display progress bars or defect updates. All job specification and results objects sent from and to the controllers and workers are also cached within Redis if they are less than N bytes (by default $N = 10$ megabytes). Worker placement and scaling are also performed by Redis. A worker node will emit an alive signal

signal and update a hash map of their current workload, the controllers will read this information periodically to make decisions to scale the cluster.

- **S3:** the job storage sub-system uses the AWS S3 protocol. As jobs progress, there are several buckets used to store various stages of a job. These buckets are:

  - Frontend jobs: currently running frontend job inputs

  - Backend jobs: currently running backend sub-job inputs

  - Frontend intermediates: intermediate data stored for certain running jobs

  - Stale frontend results: compiled results for frontend jobs (non-saved jobs)

  - Backend results: backend results of currently running jobs provided by the compute container

  - Checkpoints: last checkpoint for a running design job

  - Saved frontend jobs: store finished job inputs explicitly saved by the user

  - Saved frontend results: store compiled results from finished jobs explicitly saved by the user

  Each of these buckets also have an associated bucket marked as stale (with the exception of stale frontend results). Once an entire user job completes, all components of that job are moved into the stale bucket, with the exception of frontend job inputs and results for user saved jobs. The stale bucket has a life cycle policy of self-erasure after M days (by default $M = 45$ days). Objects less than N bytes are cached in Redis prior to being stored.

- **RDS Postgres:** Job and sub-job metadata are stored in a PostgreSQL database. For sub-jobs, the metadata stored are: ID of the affiliated job, creation, start, last update, and finish times, the type of job, the number of compute slots, and the current status of the job. For jobs, the metadata stored are: creation and finish times, the user affiliated with the job, the job type, the current status of the job, the save status of the job, any user provided comments, and the total resources used to compute the job. Metadata is used to determine job states for both the frontend and API. When a sub-job completes, the API checks all

sub-jobs for the affiliated job, and begins compilation of results if all jobs are done.

Path of a job
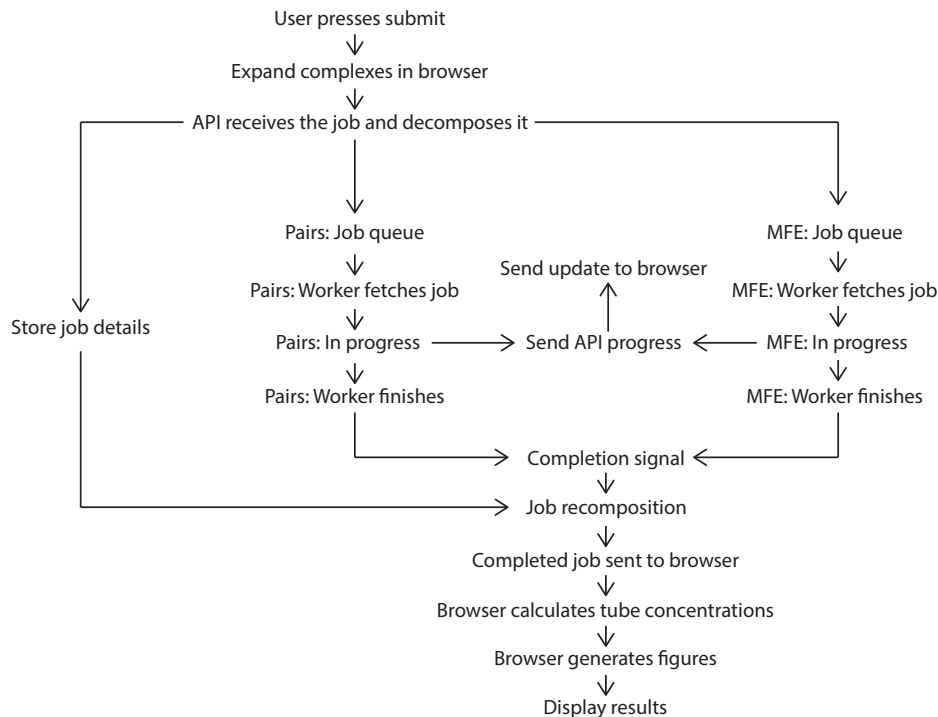
Figure 4.5: Execution graph for a single tube analysis job through the NUPACK backend.

This section will describe the path a Analysis, Design, or Utilities job will take from the user inputs to displaying results (refer to Figure 4.5 for an execution graph of a single tube analysis job). After the user submits a job, the initial phases of pre-calculations start in the user's browser on the frontend. Specifically, the expansion of strands into the test tube ensemble, subject to user-specified max complex size, and explicitly included and excluded complexes. The frontend submits this job to the controllers and receive the assigned job ID. When the job is received by the controllers, it is decomposed into sub-jobs. During this decomposition process, the estimated compute resources for the sub-job are automatically assigned and various checks are performed on the job to ensure it can be performed by NUPACK. The job and sub-job inputs are uploaded, sub-job metadata is created, and sub-jobs are submitted to the job queue.

Once a sub-job is accepted by a worker container, the worker will emit a heartbeat signal to indicate the start of a sub-job, and begin to download the job inputs and compute the job. As calculations are being done, both heartbeat and progress signals are sent. The API uses any signal sent to update the sub-job's last update time, and forwards any progress updates to the frontend. Once a sub-job finishes it sends a heartbeat signal with the finish flag. Upon receiving this signal, the API will check other sub-jobs in the database, and if they are all complete, it will pull the results for all sub-jobs, as well as the initial job inputs to create a combined final output for the job, and it stores that in the appropriate bucket.

Once a job finishes, the frontend will download the job and perform post-completion calculations. Specifically, tube concentrations are generated for analysis calculations, defects are generated for design calculations, bar plots and structure drawings are generated for both analysis and design calculations. After the post-completion steps, the job results are presented to the user.

Worker scaling

Worker containers can be distributed across multiple clusters. Each cluster is a physical or virtual subnet. Currently, two clusters are employed: one cluster at Caltech and a cluster on AWS. The Caltech cluster is fixed-size – its total core count will not change unless major additions or upgrades are made to the hardware. The AWS cluster is scalable. If there is not enough capacity for new worker containers placed on the AWS cluster, a cluster autoscaler will automatically provision a new instance for the new container. Conversely, the cluster autoscaler will periodically remove unused instances after they are inactive for some time. Both clusters are orchestrated by Kubernetes. Kubernetes was chosen for its feature set and wide acceptance within the software engineering community.

The minimum and maximum number of worker containers, worker container placement priority, and main loop delay (lower priority clusters will check for jobs less often in order to empty them first) are specified via a configuration file for each cluster. The API will always attempt to maintain O fully empty nodes (by default, $O = 3$) while meeting the above constraints. If the API decides to add a worker, it will place a pending signal in Redis, with a P minute expiry (by default, $P = 5$), and spin up the worker container using the Kubernetes API. Once a worker fully initializes, it checks for its pending signal and removes it. The API will remove empty workers if there are more than 3 empty workers by placing a deletion signal

in Redis. All workers are constantly checking for the deletion signal, and once it is received, the worker will stop accepting new jobs, finish its current jobs if there are any, send a deletion confirmation signal and shut off.

Each cluster assumes that all nodes in that cluster are uniform. As a consequence, the number of CPU cores and amount of memory per worker container is consistent for each individual cluster. The number of compute slots available for each worker node to accept jobs is proportional to the CPU and memory assigned, namely 1 job slot is 1 CPU core and 4GB of memory. The API approximates the amount of memory each job is expected to use and sends the job slots required for a particular job along with the job ID in the job queue. The main loop of a worker fetches the maximum number of jobs it can perform from a queue on Redis. The worker will then start processing those jobs, if the worker runs out of job slots before starting all the jobs fetched, it will return those unprocessed jobs back to the queue with its original rank, such that they will be prioritized for pick up by other workers.

Auxiliary functions

NUPACK offers user accounts. The user accounts will enable researchers to keep a history of their jobs. Jobs can be explicitly saved after they are started, and unsaved jobs will persist for up to 30 days. The new web app also offers download links for the NUPACK source code and Python package to all registered users.

4.5   Control plane performance

Strand	Sequence	Concentration (nM)
A	5′-GCTTGAGATGTTAGGGAGTAGTGC TCCAATCACAACGCACTACTCCCTAACATC-3′	750
B	5′-AGGGAGTAGTGCGTTGTGATTGGA AACATCTCAAGCTCCAATCACAACGCACTA-3′	750
C	5′-GTTGTGATTGGAGCTTGAGATGTT GCACTACTCCCTAACATCTCAAGCTCCAAT-3′	750

Table 4.1: Strands and concentrations used in analysis demo.

We quantified the performance of the new NUPACK control plane in three ways using variants of the Analysis demo. The Analysis demo is the equilibrium analysis of a test tube containing three strand species in solution, with a temperature melt from 35 °C to 95 °C with a 5 °C increment. Strand sequences and concentrations can be found in Table 4.1. First, we measured the time to complete jobs using the old

and new versions of the NUPACK web app. Second, we measured the effectiveness of the burstable cluster on the new web app. Third, we measured the throughput of the web app by running as many small jobs as we could in a 10-minute window.
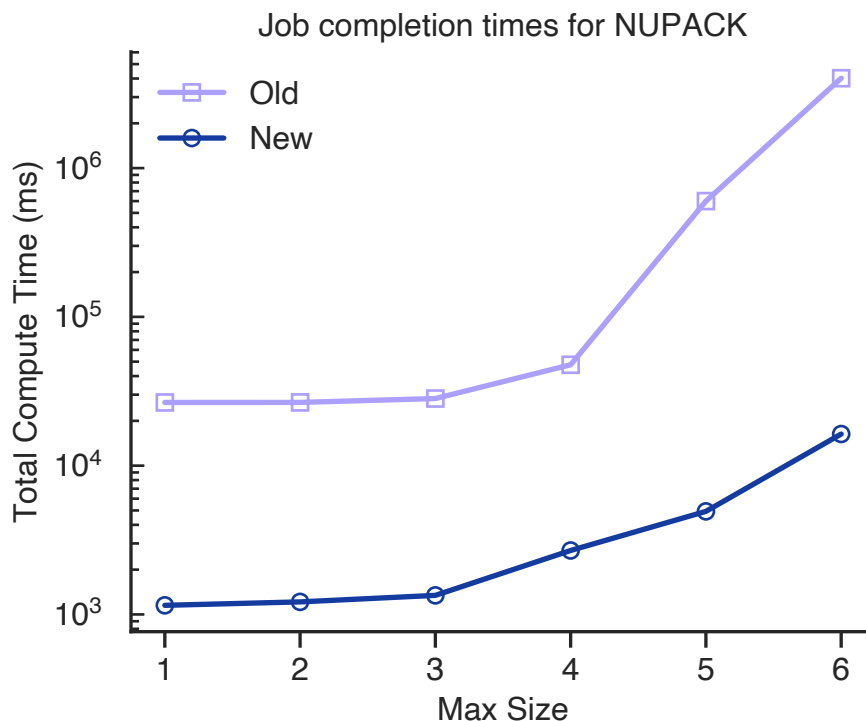

Figure 4.6: Comparison of time required to complete analysis demo with max size parameter set to 1 to 6 strands for old and new website. At low numbers of complexes, the compute should be nearly instantaneous, as such these times are representative of the web overhead (i.e. database entries, communications, etc. . .). At high number of complexes, the time in addition to the web overhead are from the calculations. The new web app is more than an order of magnitude faster for simple jobs, due to the faster web overhead; and it is more than 2 orders of magnitude faster for more complex jobs, due to the faster NUPACK4 computational backend.

To measure the time to complete jobs in the old and new versions of the NUPACK web app, we ran Analysis demo jobs with L_{max} = 1 to 6 (allow the test tube ensemble to contain all complexes of up to L_{max} strands). The smallest job in this series requires analysis of 39 complexes, while the largest requires analysis of 2925 complexes (taking account of all the temperatures in the melt). This provides a large range of job sizes to see behavioral differences for both small and large jobs. To time the jobs, we provided an email to the old web app for jobs with L_{max} set to 5 and 6 strands, and the start and finish times were taken from when the "Job Started" and "Job Finished" emails were sent. For all other tests, a screen recorder was used

to record the submission, progress, and completion of the job. The start and end times were extracted with millisecond precision from the mouse click to submit to the first frame where the complete result user interface is painted. The performance of both the new and old websites can be observed in Figure 4.6. For smaller jobs, we find the new app to be more than an order of magnitude faster. This difference is from the web overhead (i.e. database entries, communications, drawing, etc. . .) as the scientific calculations should be nearly instantaneous. For larger jobs, we find the new web app to be more than two orders of magnitude faster, due to the faster NUPACK4 computational backend.

To measure the effectiveness of the burstable cluster, we created two environments using the web app code base. Both environments use a fixed size cluster of 3 nodes with 4 cores per node. The bursting cluster is allowed to use AWS to increase its compute capacity by provisioning more 4-core nodes. The non-bursting cluster is only allowed to use the fixed cluster. We submitted 50 Analysis demo jobs with $L_{max} = 6$ over approximately one minute (translating to 1300 of each pairs and MFE sub-jobs). We observed the size of the job queue and the size of the clusters as these jobs are performed (Figure 4.7). Finishing all 50 jobs required 33 and 6 minutes for the non-bursting and bursting clusters, respectively. In the bursting cluster, we note that nodes are provisioned 3 at a time, and they generally take approximately 1 minute to provision. We determined that provisioning 3 nodes at a time is a good compromise between under-utilization and a large job queue. The ≈1-minute provisioning time is a combination of the autoscaler recognizing a need for more nodes, starting an instance, and downloading the compute code.

To measure the throughput of the web application, we scaled the number of controllers available to 10 and tried to submit (via a separate cluster) as many jobs as the control plane could accept. To accomplish this, we wrote a multi-threaded script to attempt to submit 10 jobs per second, and horizontally scaled the execution of this script across 15 nodes. To test the API controllers, the jobs are kept small to minimize compute needed. Briefly, the Analysis demo was modified use only one temperature instead of a melt, and the L_{max} was set to 1. As jobs are more frequently submitted, the access times to the controllers inevitably slows down, only the maximum number of jobs the controller can handle will be accepted. We were able to accept ≈20 jobs per second sustained across 10 minutes. Cluster utilization and node count during this process are shown in Figure 4.8.
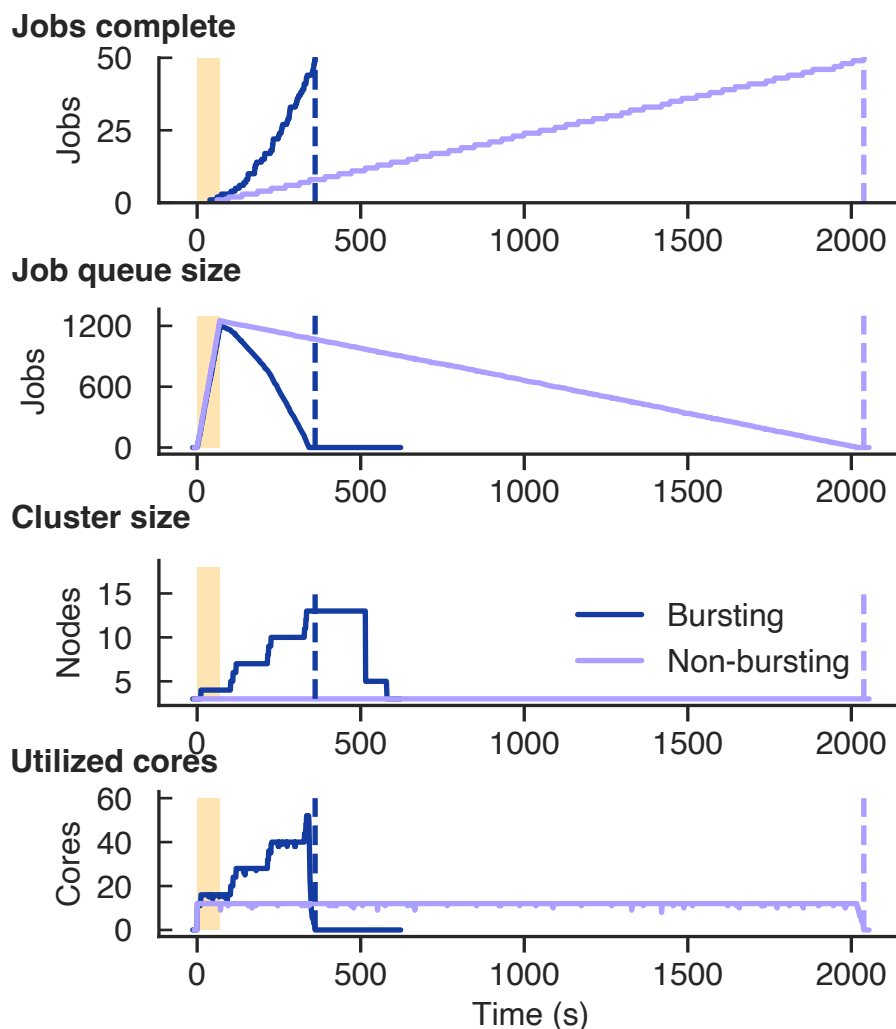
Figure 4.7: Comparison of performance to complete 50 analysis demo jobs with max size set to 6 strands for cluster with bursting to the cloud enabled and disabled. For both conditions, a base cluster of 3 nodes was used, for the bursting condition, the cluster was allowed to burst to the cloud. **Jobs complete:** number of jobs complete. **Job queue size:** number of sub-jobs waiting in the job queue. **Cluster size:** size of the cluster in number of nodes. **Utilized cores:** number of cores used to process jobs. Jobs are submitted in the time highlighted in orange, dashed lines represent full completion of all 50 jobs.

While the jobs are running, the cores utilized by these jobs fluctuates. The scientific calculations for these small jobs complete on the same time scale as fetching jobs, downloading inputs and uploading results. This causes the cluster to sometimes register low utilization, causing transient dips in cluster size as the cluster scales down. In contrast to the initial spin-up of compute nodes, these transient dips
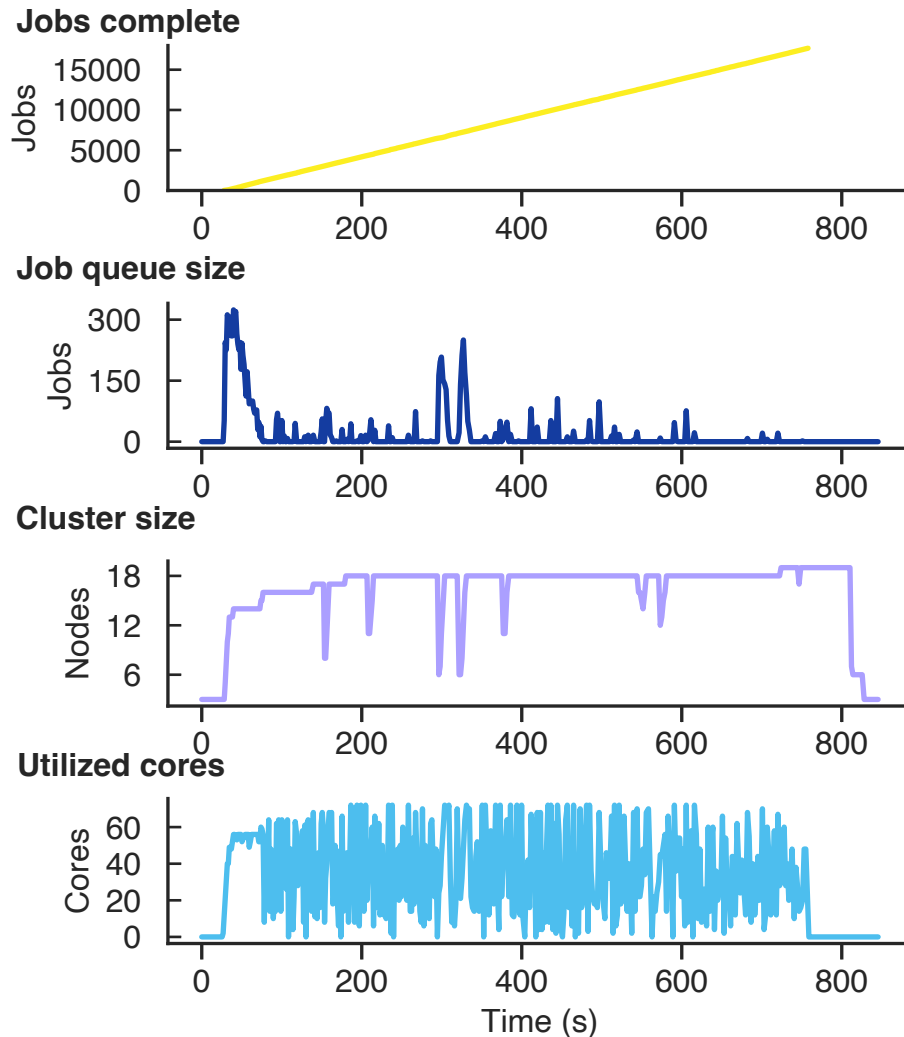
Figure 4.8: Cluster performance when the submitting as many small jobs as possible. The Analysis demo with only one temperature and $L_{max} = 1$ was used as the small job. The controllers were scaled to 10 nodes and a base cluster of 3 nodes was used; additionally the cluster was allowed to burst to AWS. A separate cluster was used to submit as many jobs as possible. **Jobs complete:** number of jobs complete, the rate is approximately 24 jobs per second. **Job queue size:** number of sub-jobs waiting in the job queue; there is an initial accumulation of jobs, but as jobs progressed, the cluster was scaled to accommodate the additional load. There is no net accumulation of jobs, demonstrating the cluster's ability to handle a large throughput of jobs. **Cluster size:** size of the cluster in number of nodes; there is an initial scaling of nodes as jobs are received. There are periods where temporary scale down occurred. However, the job load forced a scale-up shortly after. After the submissions were stopped, the nodes was scaled back to the fixed size cluster. **Utilized cores:** number of cores used to process jobs.

are almost immediately recovered as a back log of jobs build and core utilization increases. The dips recover in much less than a minute because there are two separate scalers at work. The first scaler resides in AWS infrastructure, provisioning instances when the number of nodes demanded by the controllers cannot be met with previously provisioned instances. The second scaler resides in the NUPACK controllers, changing the number of nodes demanded. The AWS scaler is at a lower level than the NUPACK scaler. When NUPACK removes a node, it does not remove the underlying instance, and as such, that instance is available to host a new NUPACK compute node, should the need arise. Instances will scale down after they are unused for a number of minutes.

C h a p t e r 5

CONCLUSIONS

This thesis touches on a variety of projects focused on nucleic acid nanotechnologies. It presents a step towards the application of nucleic acids in closed-loop synthetic genetic circuits and molecular diagnostics. It also improves the tools used by the nucleic acid research community to analyze and design nucleic acid systems. Each of these project has several future directions to pursue.

In the field of synthetic biology, researchers are always looking for more scalable and orthogonal ways to build synthetic genetic circuits. In this thesis, we show the potential of cgRNAs to fulfill this role. To demonstrate the orthogonality and feedback functionality of cgRNAs, we show a simple feedback circuit built with cgRNAs. This circuit has three distinct open- or closed-loop configurations with varying transfer curves, and two of the same circuits can function in the same cell at the same time without interference with each other. The versatility of cgRNA circuit construction was demonstrated by combining the components of the previous two simple circuits and making a larger toggle switch. The toggle switch held the respective toggled state in the presence and absence of the initial trigger. This work demonstrate cgRNAs as a valuable component of functional scalable synthetic genetic circuits in cells.

The COVID-19 pandemic has uncovered the weakness of the current regime of molecular diagnostics for novel pathogens. The weakness of current commercial rapid antigen tests are: 1) the large time gap between genomic sequencing of the pathogen and the development and screening of effective antibody pairs; and 2) the high limit of detection. The large time gap leads to non-continuous laboratory testing and potentially supply shortages in the initial stages of an emerging pathogen. The sensitivity gap leads to false-negative tests, allowing infectious individuals to roam free and infect others. We demonstrated an HCR-amplified test that detects the genomic RNA in the virus rather than a protein. This overcomes both weaknesses of a traditional rapid antigen test. First, due to the programmable nature of nucleic-acid based testing technologies, probes can be designed, manufactured, and purified in days instead of months from the date of genomic sequencing of a novel pathogen. Second, in the test described in this thesis, we were able to decrease limit

of detection by 2.5- to 100-fold compared to commercial antigen tests. We recognize the drawbacks of the HCR-amplified viral RNA test, specifically a 15-minute warming step and longer overall test time. However, despite these shortcomings, it still represents a potential solution and exciting research direction both for testing existing pathogens with better sensitivity and rapidly responding to future emergent pathogens.

Both cgRNAs and HCR systems are designed by NUPACK. NUPACK is a suite of tools for nucleic acid design and analysis. The final project in this thesis describes a new NUPACK web app rearchitected and rebuilt from the ground up. Since the last version of the NUPACK web app was released, there were new developments in the backend recursions [92] and design formulation [8]. While sustaining rapid growth and usage from the research community, the software and hardware used to develop the previous NUPACK web app fell into obsolesce. As a result, the older NUPACK web app is slow and the queue is often full, requiring manual reboots or queue clears. This thesis described the user interface and technical details of the new NUPACK web implementation. This new version includes allowing the control plane and clusters to burst dynamically to job demand, updating the compute backend [92], and implementing various new user interface features. This further development of NUPACK will enable nucleic acid designers to create larger and more complex nucleic acid structures, pathways, and systems.

5.1 Future directions

In all three projects, there are additional long- and short-term future directions of research. These directions will be explored in the following section.

Using conditional guide RNAs for scalable synthetic feedback circuits

In the field of synthetic genetic circuits, there are many future directions that may take advantage of cgRNA circuits. These future directions include varied circuit motifs, different circuit settings, further cgRNA developments, and the incorporation of computational design and improved validation techniques.

There are many more circuit motifs that can be tested in future derivatives of the studies described in this thesis. Additional functional motifs such as feedforward loops, logical components, or memory could possibly be implemented using cgRNAs in living cells.

Circuits can be built and tested in various non-bacterial cell types, such as yeast, mammalian cells, or embryos, thanks to the portability of cgRNAs and the CRISPR system [66, 72]. Furthermore, inducer synthases, such as AHL synthases, enable cgRNA circuits to be integrated into multi-cellular and inter-species systems. Examples of these circuits include multi-species consortia of bacteria that can spatiotemporally regulate themselves in response to external stimuli, or bacteria with a synthetic circuit that could interface with more complex organisms such as the root systems of plants.

Further cgRNA technology development could lead to larger, more scalable, and more versatile circuits. New modes of CRISPR systems can enable more functions in the cell. For example, Cas13a can be used to cleave formed mRNAs, unmodified Cas9 can be used to cleave dsDNA, and various activators exist to turn on transcription from an inactive promoter. These new variants of effector proteins can be tested with current and new cgRNAs and circuits. Synthetic circuits can also be integrated into native cell biology by using mRNAs as cgRNA triggers. These developments will present a more versatile repertoire of tools for larger scale circuits.

Additionally, in order to enable larger circuits, we need larger libraries of orthogonal cgRNAs. The largest cgRNA library published in the Pierce lab is currently 4 [104]. Recent developments in rational design allows researchers to design multiple orthogonal systems of nucleic acid pathways at once through a new multi-state designer [8]. With the new designer features, such as weighting of domains, energy constraints, and the advent of a much faster core code, larger sets of orthogonal cgRNAs can be designed at a faster pace. Combined with commercial developments in DNA and plasmid purchasing options, we can more easily create and validate a large library of orthogonal cgRNAs and triggers. This library of cgRNAs will be crucial for any future circuit development work.

Currently, circuit construction requires significant fine tuning of the circuit's components and physical plasmids. This process is often tedious and will be prohibitive in terms of cost and labor for larger circuits. To alleviate this, we can extract computation models and parameters of cgRNA function through system ID and a systematic series of experiments. In the future, we hope to achieve sufficient predictive power in our simulations and/or screening techniques such that we can build and iterate our circuits in shorter time frames to enable the design and validation of larger circuit architectures.

With these larger orthogonal sets of cgRNAs, more versatile libraries of effector Cas proteins, a large set of biological settings, and accurate computational models, cgRNAs could provide an ideal platform for large-scale programmable synthetic biology.

HCR Lateral Flow Assays for Amplified At-Home SARS-CoV-2 Testing

In the field of molecular diagnostics, there are many improvements that can be made to the prototype device described in this thesis. Improvements can be functional, such as: decreasing the limit of detection and enabling multiplex detection of multiple targets at the same time. Improvements can also be for user convenience, such as: eliminating the 15-minute heat step and manufacturing a device to easily assemble and perform the test. Further characterization could elucidate additional important properties of the test such as shelf life and bio-orthogonality to other drugs, viruses or analytes.

The limit of detection for the prototype device can be improved. In this thesis, the ratio of amplification between using a H1 hairpin only and both H1 and H2 hairpins is $\approx$10. However in previous studies [21], an amplification ratio up to 450 can be observed. We have two hypotheses for this discrepancy. First, the experiment duration, concentration of probes, concentration of hairpins, and reagent conditions are vastly different than previous works with HCR amplification, and further optimization could yield higher hairpin polymer lengths. Second, if hairpins form long polymers, then steric hindrance of the large carbon black particles or the anti-Dig can occlude some haptens on the polymer chain from binding to carbon black.

The programmability of a nucleic acid-based test enables multiplexed detection of different analytes of interest on a single test. Different detection targets can use different probe sets or hairpins mixed in one-pot. The current foreseeable challenge to this approach lies in using an anti-DNA:RNA hybrid antibody as the means to immobilize the analyte at the test line. To create distinctive test lines, the immobilized capture probes will need to distinguish between targets. There are two possible ways to overcome this: 1) using modified nucleic acids with faster kinetics to capture genome, and 2) the modification of DNA initator probes or use of specialized capture probes to include orthogonal haptens that can interact with immobilized anti-hapten antibodies.

In terms of usability and convenience, the heat-step would be the last potential roadblock to deploying this test in an at-home setting. Removing the heat-step requires an automated chemical or physical process that will allow the virus to release the genomic RNA, but must not inhibit DNA probe to RNA genome base pairing. Preliminary studies were done with various chemical denaturants, however signal was not observed in these studies. Further studies can be conducted using additional chemical conditions, incorporation of a wash channel, or through a physical process to break and remove viral envelopes. After the heat step is eliminated, a 1-step test will be possible by drying down various reagents onto conjugate pads.

On the topic of commercialization, additional studies, such as the shelf life of various reagents used and the orthogonality of the test to drugs and other viruses, will need to be performed on the finished device prior to commercialization. Furthermore, clinical data collection and presentation will most likely be necessary to commercialize.

NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems

We rearchitected and rewrote the NUPACK web app from the ground up to enable molecular biologists and nucleic acid designers to analyze and design larger and more complex systems of nucleic acid structures, devices, and systems. Though great improvements in speed and availability are currently achieved, there are many features that could further increase the utility of this indispensable tool: features such as tighter integration with Python-based web notebooks, and allowing users to submit batch jobs will enable the power user to rapidly prototype and screen many concurrent systems *in silico*. Further scientific approaches to nucleic acid analysis, such as kinetic trajectory analysis of nucleic acid binding, in contrast to current thermodynamic equilibrium analyses, may also be useful for the community at large. As the custodians of the tool, we will also need to solidify our archival and data retention policies of user ran jobs and data. As we move out of the initial release of the new NUPACK web platform, we will be moving into a new phase of continuous development.

Bibliography

[1]   M. H. Hanewich-Hollatz, Z. Chen, L. M. Hochrein, J. Huang, and N. A. Pierce. Conditional guide RNAs: Programmable conditional regulation of CRISPR/Cas function in bacterial and mammalian cells via dynamic RNA nanotechnology. *ACS Central Science*, 5(7):1241–1249, 2019.

[2]   M. J. Serra and D. H. Turner. Predicting thermodynamic properties of RNA. *Methods in Enzymology*, 259:242–261, 1995.

[3]   D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology*, 288:911–940, 1999.

[4]   R. M. Dirks and N. A. Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *Journal of Computational Chemistry*, 24(13):1664–1677, 2003. ISSN 1096-987X. doi: 10.1002/jcc. 10296. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.10296`.

[5]   R. M. Dirks, J. S. Bois, J. M. Schaeffer, E. Winfree, and N. A. Pierce. Thermodynamic analysis of interacting nucleic acid strands. *SIAM Rev.*, 49 (1):65–88, 2007.

[6]   J. N. Zadeh, B. R. Wolfe, and N. A. Pierce. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry*, 32(3):439–452, 2011. ISSN 1096-987X. doi: 10.1002/jcc.21633. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21633`.

[7]   B. R. Wolfe and N. A. Pierce. Sequence design for a test tube of interacting nucleic acid strands. *ACS Synthetic Biology*, 4(10):1086–1100, October 2015.

[8]   B. R. Wolfe, N. J. Porubsky, J. N. Zadeh, R. M. Dirks, and N. A. Pierce. Constrained multistate sequence design for nucleic acid reaction pathway engineering. *Journal of the American Chemical Society*, 139(8):3134–3144, March 2017.

[9]   P. W. K. Rothemund. Folding DNA to create nanoscale shapes and patterns. *Nature*, 440(7082):297–302, 2006.

[10] G. Tikhomirov, P. Petersen, and L. Qian. Fractal assembly of micrometre-scale DNA origami arrays with arbitrary patterns. *Nature*, 552(7683):67–71, December 2017. ISSN 1476-4687. doi: 10.1038/nature24655. URL `https://www.nature.com/articles/nature24655`.

[11] L. M. Hochrein, M. Schwarzkopf, M. Shahgholi, P. Yin, and N. 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.

[12] L. M. Hochrein, T. J. Ge, M. Schwarzkopf, and N. A. Pierce. Signal transduction in human cell lysate via dynamic RNA nanotechnology. *ACS Synthetic Biology*, 7(12):2796–2802, 2018.

[13] J. Chappell, K. E. Watters, M. K. Takahashi, and J. B. Lucks. A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology*, 28:47–56, 2015.

[14] A. A. Green, J. Kim, D. Ma, P. A. Silver, J. J. Collins, and P. Yin. Complex cellular logic computation using ribocomputing devices. *Nature*, 548(7665): 117–121, August 2017. ISSN 1476-4687. doi: 10.1038/nature23271. URL `https://www.nature.com/articles/nature23271`.

[15] R. M. Dirks and N. A. Pierce. Triggered amplification by hybridization chain reaction. *Proceedings of the National academy of Sciences*, 101(43): 15275–15278, 2004.

[16] H. M. T. Choi, J. Y. Chang, L. A. Trinh, J. E. Padilla, S. E. Fraser, and N. A. Pierce. Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nature Biotechnology*, 28(11):1208–12, 2010. ISSN 1546-1696 (Electronic)1087-0156 (Linking). doi: 10.1038/nbt.1692.

[17] H. M. T. Choi, V. A. Beck, and N. A. Pierce. Next-generation in situ hybridization chain reaction: Higher gain, lower cost, greater durability. *ACS Nano*, 8(5):4284–4294, 2014. doi: 10.1021/nn405717p.

[18] S. Shah, E. Lubeck, M. Schwarzkopf, T.-F. He, A. Greenbaum, C. H. Sohn, A. Lignell, H. M. T. Choi, V. Gradinaru, N. A. Pierce, and L. Cai. Single-molecule RNA detection at depth via hybridization chain reaction and tissue hydrogel embedding and clearing. *Development*, 143:2862–2867, 2016. doi: 10.1242/dev.138560.

[19] H. M. T. Choi, C. R. Calvert, N. Husain, D. Huss, J. C. Barsi, B. E. Deverman, R. C. Hunter, M. Kato, S. M. Lee, A. C. T. Abelin, A. Z. Rosenthal, O. S. Akbari, Y. Li, B. A. Hay, P. W. Sternberg, P. H. Patterson, E. H. Davidson, S. K. Mazmanian, D. A. Prober, M. van de Rijn, J. R. Leadbetter, D. K. Newman, C. Readhead, M. E. Bronner, B. Wold, R. Lansford, T. Sauka-Spengler, S. E. Fraser, and N. A. Pierce. Mapping a multiplexed zoo of mRNA expression. *Development*, 143:3632–3637, 2016. doi: 10.1242/dev.140137.

[20] H. M. T. Choi, M. Schwarzkopf, M. E. Fornace, A. Acharya, G. Artavanis, J. Stegmaier, A. Cunha, and N. A. Pierce. Third-generation in situ hybridization chain reaction: Multiplexed, quantitative, sensitive, versatile, robust. *Development*, 145:dev165753, 2018. doi: 10.1242/dev.165753.

[21] M. Schwarzkopf, M. C. Liu, S. J. Schulte, R. Ives, N. Husain, H. M. T. Choi, and N. A. Pierce. Hybridization chain reaction enables a unified approach to multiplexed, quantitative, high-resolution immunohistochemistry and *in situ*

hybridization. *Development*, 148(22):dev199847, November 2021. ISSN 0950-1991, 1477-9129. doi: 10.1242/dev.199847.

[22] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, June 1961. ISSN 0022-2836. doi: 10.1016/S0022-2836(61)80072-7. URL `https://www.sciencedirect.com/science/article/pii/S0022283661800727`.

[23] S. Ogden, D. Haggerty, C. M. Stoner, D. Kolodrubetz, and R. Schleif. The Escherichia coli L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proceedings of the National Academy of Sciences*, 77(6):3346–3350, June 1980. doi: 10.1073/pnas.77.6.3346. URL `https://www.pnas.org/doi/10.1073/pnas.77.6.3346`.

[24] W. Hinrichs, C. Kisker, M. Düvel, A. Müller, K. Tovar, W. Hillen, and W. Saenger. Structure of the tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science*, 264(5157):418–420, April 1994. doi: 10.1126/science.8153629. URL `https://www.science.org/doi/10.1126/science.8153629`.

[25] M. B. Miller and B. L. Bassler. Quorum sensing in bacteria. *Annual Review of Microbiology*, 55(1):165–199, 2001. doi: 10.1146/annurev.micro.55.1.165. URL `https://doi.org/10.1146/annurev.micro.55.1.165`.

[26] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, January 1999. ISSN 1476-4687. doi: 10.1038/16483. URL `https://www.nature.com/articles/16483`.

[27] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, 31 (1):64–68, May 2002. ISSN 1546-1718. doi: 10.1038/ng881. URL `https://www.nature.com/articles/ng881z`.

[28] O. S. Venturelli, H. El-Samad, and R. M. Murray. Synergistic dual positive feedback loops established by molecular sequestration generate robust bimodal response. *Proceedings of the National Academy of Sciences*, 109 (48):E3324–E3333, November 2012. doi: 10.1073/pnas.1211902109. URL `https://www.pnas.org/doi/10.1073/pnas.1211902109`.

[29] H. El-samad, J. P. Goff, and M. Khammash. Calcium homeostasis and parturient hypocalcemia: An integral feedback perspective. *Journal of Theoretical Biology*, 214(1):17–29, January 2002. ISSN 0022-5193. doi: 10.1006/jtbi.2001.2422. URL `https://www.sciencedirect.com/science/article/pii/S0022519301924221`.

[30] G. D. Snell and G. F. Higgins. Alleles at the histocompatibility-2 locus in the mouse as determined by tumor transplantation. *Genetics*, 36(3):306–310, May 1951. ISSN 1943-2631. doi: 10.1093/genetics/36.3.306. URL `https://doi.org/10.1093/genetics/36.3.306`.

[31] C.-H. Lecellier, P. Dunoyer, K. Arar, J. Lehmann-Che, S. Eyquem, C. Himber, A. Saïb, and O. Voinnet. A cellular microrna mediates antiviral defense in human cells. *Science*, 308(5721):557–560, April 2005. doi: 10.1126/science.1108784. URL `https://www.science.org/doi/10.1126/science.1108784`.

[32] P. Horvath and R. Barrangou. CRISPR/Cas, the immune system of bacteria and archaea. *Science*, 327(5962):167–170, January 2010. doi: 10.1126/science.1179555. URL `https://www.science.org/doi/10.1126/science.1179555`.

[33] E. C. Lai. Notch signaling: Control of cell communication and cell fate. *Development*, 131(5):965–973, March 2004. ISSN 0950-1991. doi: 10.1242/dev.01074. URL `https://doi.org/10.1242/dev.01074`.

[34] W. Driever and C. Nüsslein-Volhard. The bicoid protein determines position in the Drosophila embryo in a concentration-dependent manner. *Cell*, 54(1):95–104, July 1988. ISSN 0092-8674. doi: 10.1016/0092-8674(88)90183-3. URL `https://www.sciencedirect.com/science/article/pii/0092867488901833`.

[35] Y. Zhao. Auxin biosynthesis and its role in plant development. *Annual Review of Plant Biology*, 61(1):49–64, 2010. doi: 10.1146/annurev-arplant-042809-112308. URL `https://doi.org/10.1146/annurev-arplant-042809-112308`.

[36] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, January 2000.

[37] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000. ISSN 1476-4687. doi: 10.1038/35002131. URL `https://www.nature.com/articles/35002131`.

[38] F. Wang, J. Elbaz, R. Orbach, N. Magen, and I. Willner. Amplified analysis of DNA by the autonomous assembly of polymers consisting of DNAzyme wires. *Journal of the American Chemical Society*, 133(43):17149–17151, 2011. ISSN 0002-7863. doi: 10.1021/Ja2076789.

[39] V. Singh. Recent advances and opportunities in synthetic logic gates engineering in living cells. *Systems and Synthetic Biology*, 8(4):271–282, December 2014. ISSN 1872-5325. doi: 10.1007/s11693-014-9154-6. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4571725/`.

[40] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins. Synthetic gene networks that count. *Science*, 324(5931):1199–1202, May 2009. doi: 10.1126/science.1172005. URL `https://www.science.org/doi/10.1126/science.1172005`.

[41] S. Basu, R. Mehreja, S. Thiberge, M.-T. Chen, and R. Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *Proceedings of the National Academy of Sciences*, 101(17):6355–6360, April 2004. doi: 10.1073/pnas.0307571101. URL `https://www.pnas.org/doi/full/10.1073/pnas.0307571101`.

[42] F. Annunziata, A. Matyjaszkiewicz, G. Fiore, C. S. Grierson, L. Marucci, M. di Bernardo, and N. J. Savery. An orthogonal multi-input integration system to control gene expression in Escherichia coli. *ACS Synthetic Biology*, 6(10):1816–1824, October 2017. doi: 10.1021/acssynbio.7b00109. URL `https://doi.org/10.1021/acssynbio.7b00109`.

[43] J. Bonnet, P. Subsoontorn, and D. Endy. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proceedings of the National Academy of Sciences*, 109(23):8884–8889, June 2012. doi: 10.1073/pnas.1202344109. URL `https://www.pnas.org/doi/10.1073/pnas.1202344109`.

[44] V. Hsiao, Y. Hori, P. W. K. Rothemund, and R. M. Murray. A population-based temporal logic gate for timing and recording chemical events. *Molecular Systems Biology*, 12(5):869, May 2016. ISSN 1744-4292. doi: 10.15252/msb.20156663. URL `https://www.embopress.org/doi/full/10.15252/msb.20156663`.

[45] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037): 1130–1134, April 2005. ISSN 1476-4687. doi: 10.1038/nature03461. URL `https://www.nature.com/articles/nature03461`.

[46] L. Potvin-Trottier, N. D. Lord, G. Vinnicombe, and J. Paulsson. Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, 538(7626):514–517, October 2016. ISSN 1476-4687. doi: 10.1038/nature19841. URL `https://www.nature.com/articles/nature19841`.

[47] Z. Gao, E. Herrera-Carrillo, and B. Berkhout. Delineation of the exact transcription termination signal for type 3 polymerase III. *Molecular Therapy - Nucleic Acids*, 10:36–44, 2018.

[48] V. Hsiao, A. Swaminathan, and R. M. Murray. Control theory for synthetic biology: Recent advances in system characterization, control design, and controller implementation for synthetic biology. *IEEE Control Systems Magazine*, 38(3):32–62, June 2018. ISSN 1941-000X. doi: 10.1109/MCS.2018.2810459.

[49] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, 323(5):785–793, November 2002. ISSN 0022-2836. doi: 10.1016/S0022-2836(02)00994-4. URL `https://www.sciencedirect.com/science/article/pii/S0022283602009944`.

[50] L. Goentoro, O. Shoval, M. W. Kirschner, and U. Alon. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular Cell*, 36(5):894–899, December 2009. ISSN 1097-4164. doi: 10.1016/j.molcel.2009.11.018.

[51] R. D. McCardell, S. Huang, L. N. Green, and R. M. Murray. Control of bacterial population density with population feedback and molecular sequestration, November 2017. URL `https://www.biorxiv.org/content/10.1101/225045v1`.

[52] V. Hsiao, E. L. C. de los Santos, W. R. Whitaker, J. E. Dueber, and R. M. Murray. Design and implementation of a biomolecular concentration tracker. *ACS Synthetic Biology*, 4(2):150–161, February 2015. doi: 10.1021/sb500024b. URL `https://doi.org/10.1021/sb500024b`.

[53] J. A. N. Brophy and C. A. Voigt. Principles of genetic circuit design. *Nature Methods*, 11(5):508–520, May 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2926. URL `https://www.nature.com/articles/nmeth.2926`.

[54] P. E. M. Purnick and R. Weiss. The second wave of synthetic biology: From modules to systems. *Nature Reviews Molecular Cell Biology*, 10(6):410–422, June 2009. ISSN 1471-0080. doi: 10.1038/nrm2698. URL `https://www.nature.com/articles/nrm2698`.

[55] E. Leonard, D. Nielsen, K. Solomon, and K. J. Prather. Engineering microbes with synthetic biology frameworks. *Trends in Biotechnology*, 26(12):674–681, December 2008. ISSN 0167-7799. doi: 10.1016/j.tibtech.2008.08.003.

[56] V. Singh. Recent advancements in synthetic biology: Current status and challenges. *Gene*, 535(1):1–11, February 2014. ISSN 0378-1119. doi: 10.1016/j.gene.2013.11.025. URL `https://www.sciencedirect.com/science/article/pii/S0378111913015515`.

[57] D. E. Cameron, C. J. Bashor, and J. J. Collins. A brief history of synthetic biology. *Nature Reviews Microbiology*, 12(5):381–390, 2014. ISSN 1740-1526. doi: 10.1038/nrmicro3239.

[58] A. S. Khalil and J. J. Collins. Synthetic biology: Applications come of age. *Nature Reviews Genetics*, 11(5):367–379, May 2010. ISSN 1471-0064. doi: 10.1038/nrg2775. URL `https://www.nature.com/articles/nrg2775`.

[59] A.-A. Baetica, Y. P. Leong, N. Olsman, and R. M. Murray. Design guidelines for sequestration feedback networks, October 2018. URL `https://www.biorxiv.org/content/10.1101/455493v1`.

[60] T. Shopera, W. R. Henson, and T. S. Moon. Dynamics of sequestration-based gene regulatory cascades. *Nucleic Acids Research*, 45(12):7515–7526, July 2017. ISSN 1362-4962. doi: 10.1093/nar/gkx465.

[61] A. J. Meyer, T. H. Segall-Shapiro, E. Glassey, J. Zhang, and C. A. Voigt. Escherichia coli "Marionette" strains with 12 highly optimized small-molecule sensors. *Nature Chemical Biology*, 15(2):196–204, February 2019. ISSN 1552-4469. doi: 10.1038/s41589-018-0168-3. URL `https://www.nature.com/articles/s41589-018-0168-3`.

[62] L. Qian and E. Winfree. Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034):1196, June 2011. doi: 10.1126/science.1200520.

[63] H.-H. Huang, Y. Qian, and D. Del Vecchio. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nature Communications*, 9(1):5415, December 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-07899-z. URL `https://www.nature.com/articles/s41467-018-07899-z`.

[64] C. L. Kelly, A. W. K. Harris, H. Steel, E. J. Hancock, J. T. Heap, and A. Papachristodoulou. Synthetic negative feedback circuits using engineered small RNAs. *Nucleic Acids Research*, 46(18):9875–9889, October 2018. ISSN 0305-1048. doi: 10.1093/nar/gky828. URL `https://doi.org/10.1093/nar/gky828`.

[65] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier. A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337(6096):816–821, August 2012. doi: 10.1126/science.1225829. URL `https://www.science.org/doi/10.1126/science.1225829`.

[66] L. Cong, F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini, and F. Zhang. Multiplex genome engineering using CRISPR/Cas systems. *Science*, pages 819–823, January 2013.

[67] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, and L. A. Marraffini. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, 41(15):7429–7437, August 2013. ISSN 1362-4962. doi: 10.1093/nar/gkt520.

[68] A. Chavez, M. Tuttle, B. W. Pruitt, B. Ewen-Campen, R. Chari, D. Ter-Ovanesyan, S. J. Haque, R. J. Cecchi, E. J. K. Kowal, J. Buchthal, B. E. Housden, N. Perrimon, J. J. Collins, and G. Church. Comparison of Cas9

activators in multiple species. *Nature Methods*, 13(7):563–567, July 2016. ISSN 1548-7105. doi: 10.1038/nmeth.3871. URL https://www.nature.com/articles/nmeth.3871.

[69] C. Dong, J. Fontana, A. Patel, J. M. Carothers, and J. G. Zalatan. Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications*, 9(1):2489, June 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-04901-6.

[70] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim. Repurposing CRISPR as an RNA-Guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–1183, 2013. ISSN 0092–8674. doi: 10.1016/j.cell.2013.02.022.

[71] J. Kim, Y. Zhou, P. D. Carlson, M. Teichmann, S. Chaudhary, F. C. Simmel, P. A. Silver, J. J. Collins, J. B. Lucks, P. Yin, and A. A. Green. De novo-designed translation-repressing riboregulators for multi-input cellular logic. *Nature Chemical Biology*, 15(12):1173–1182, December 2019. ISSN 1552-4469. doi: 10.1038/s41589-019-0388-1.

[72] M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, and E. Klavins. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nature Communications*, 8(1):15459, May 2017. ISSN 2041-1723. doi: 10.1038/ncomms15459.

[73] L. Nissim, S. D. Perli, A. Fridkin, P. Perez-Pinera, and T. K. Lu. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, 2014.

[74] S. Kiani, J. Beal, M. R. Ebrahimkhani, J. Huh, R. N. Hall, Z. Xie, Y. Li, and R. Weiss. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nature Methods*, 11(7):723–726, July 2014. ISSN 1548-7105. doi: 10.1038/nmeth.2969. URL https://www.nature.com/articles/nmeth.2969.

[75] F. Ceroni, A. Boo, S. Furini, T. E. Gorochowski, O. Borkowski, Y. N. Ladak, A. R. Awan, C. Gilbert, G.-B. Stan, and T. Ellis. Burden-driven feedback control of gene expression. *Nature Methods*, 15(5):387–393, May 2018. ISSN 1548-7105. doi: 10.1038/nmeth.4635. URL https://www.nature.com/articles/nmeth.4635.

[76] Y. Nihongaki, T. Otabe, and M. Sato. Emerging approaches for spatiotemporal control of targeted genome with inducible CRISPR-Cas9. *Analytical Chemistry*, 90(1):429–439, January 2018.

[77] B. J. Aubrey, G. L. Kelly, A. J. Kueh, M. S. Brennan, L. O'Connor, L. Milla, S. Wilcox, L. Tai, A. Strasser, and M. J. Herold. An inducible lentiviral guide RNA platform enables the identification of tumor-essential genes and tumor-promoting mutations in vivo. *Cell Reports*, 10(8):1422–1432, 2015.

[78] K.-H. Siu and W. Chen. Riboregulated toehold-gated gRNA for programmable CRISPR–Cas9 function. *Nature Chemical Biology*, 15:217–220, 2019.

[79] L. Oesinghaus and F. C. Simmel. Switching the activity of Cas12a using guide RNA strand displacement circuits. *Nature Communications*, 10(1): 1–11, 2019. ISSN 2041-1723. doi: 10.1038/s41467-019-09953-w.

[80] M. Jin, N. Garreau de Loubresse, Y. Kim, J. Kim, and P. Yin. Programmable CRISPR-Cas repression, activation, and computation with sequence-independent targets and triggers. *ACS Synthetic Biology*, 8(7): 1583–1589, July 2019. ISSN 2161-5063, 2161-5063. doi: 10.1021/acssynbio.9b00141.

[81] CDC. CDC Museum COVID-19 Timeline, August 2022. URL `https://www.cdc.gov/museum/timeline/covid19.html`.

[82] B. Abdalhamid, C. R. Bilder, E. L. McCutchen, S. H. Hinrichs, S. A. Koepsell, and P. C. Iwen. Assessment of specimen pooling to conserve SARS CoV-2 testing resources. *American Journal of Clinical Pathology*, 153(6):715–718, May 2020. ISSN 0002-9173. doi: 10.1093/ajcp/aqaa064. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7188150/`.

[83] Office of the Commissioner. Coronavirus (COVID-19) update: FDA authorizes first antigen test to help in the rapid detection of the virus that causes COVID-19 in patients, May 2020. URL `https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-antigen-test-help-rapid-detection-virus-causes`. Publisher: FDA.

[84] Upping the ante on COVID-19 antigen testing, 2020. URL `https://www.abbott.com/corpnewsroom/diagnostics-testing/upping-the-ante-on-COVID-19-antigen-testing.html`.

[85] G. Caruana, A. Croxatto, E. Kampouri, A. Kritikos, O. Opota, M. Foerster, R. Brouillet, L. Senn, R. Lienhard, A. Egli, Giuseppe Pantaleo, P.-N. Carron, and G. Greub. Implementing SARS-CoV-2 rapid antigen testing in the emergency ward of a swiss university hospital: The IN-CREASE study. *Microorganisms*, 9(4):798, 2021. ISSN 2076-2607. doi: 10.3390/microorganisms9040798.

[86] M. Wölfl-Duchek, F. Bergmann, A. Jorda, M. Weber, M. Müller, T. Seitz, A. Zoufaly, R. Strassl, M. Zeitlinger, H. Herkner, H. Schnidar, K. Anderle, and U. Derhaschnig. Sensitivity and specificity of SARS-CoV-2 rapid antigen detection tests using oral, anterior nasal, and nasopharyngeal swabs: A diagnostic accuracy study. *Microbiology Spectrum*, 10(1):e02029–21, February 2022. ISSN 2165-0497. doi: 10.1128/spectrum.02029-21.

[87] M. J. Flynn, O. Snitser, J. Flynn, S. Green, I. Yelin, M. Szwarcwort-Cohen, R. Kishony, and M. B. Elowitz. A simple direct RT-LAMP SARS-CoV-2 saliva diagnostic, November 2020. URL `https://www.medrxiv.org/content/10.1101/2020.11.19.20234948v1`.

[88] V. L. Dao Thi, K. Herbst, K. Boerner, M. Meurer, L. P. M. Kremer, D. Kirrmaier, A. Freistaedter, D. Papagiannidis, C. Galmozzi, M. L. Stanifer, S. Boulant, S. Klein, P. Chlanda, D. Khalid, I. Barreto Miranda, P. Schnitzler, H.-G. Kräusslich, M. Knop, and S. Anders. A colorimetric RT-LAMP assay and LAMP-sequencing for detecting SARS-CoV-2 RNA in clinical samples. *Science Translational Medicine*, 12(556):eabc7075, August 2020. doi: 10.1126/scitranslmed.abc7075. URL `https://www.science.org/doi/10.1126/scitranslmed.abc7075`.

[89] T. Notomi, H. Okayama, H. Masubuchi, T. Yonekawa, K. Watanabe, N. Amino, and T. Hase. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, 28(12):e63, June 2000. ISSN 0305-1048. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC102748/`.

[90] J. S. Gootenberg, O. O. Abudayyeh, J. W. Lee, P. Essletzbichler, A. J. Dy, J. Joung, V. Verdine, N. Donghia, N. M. Daringer, C. A. Freije, C. Myhrvold, R. P. Bhattacharyya, J. Livny, A. Regev, E. V. Koonin, D. T. Hung, P. C. Sabeti, J. J. Collins, and F. Zhang. Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, April 2017. doi: 10.1126/science.aam9321. URL `https://www.science.org/doi/10.1126/science.aam9321`.

[91] H. de Puig, R. A. Lee, D. Najjar, X. Tan, L. R. Soenksen, N. M. Angenent-Mari, N. M. Donghia, N. E. Weckman, A. Ory, C. F. Ng, P. Q. Nguyen, A. S. Mao, T. C. Ferrante, G. Lansberry, H. Sallum, J. Niemi, and J. J. Collins. Minimally instrumented SHERLOCK (miSHERLOCK) for CRISPR-based point-of-care diagnosis of SARS-CoV-2 and emerging variants. *Science Advances*, 7(32):eabh2944, August 2021. doi: 10.1126/sciadv.abh2944. URL `https://www.science.org/doi/10.1126/sciadv.abh2944`.

[92] M. E. Fornace, N. J. Porubsky, and N. A. Pierce. A unified dynamic programming framework for the analysis of interacting nucleic acid strands: Enhanced models, scalability, and speed. *ACS Synthetic Biology*, 9(10): 2665–2678, 2020. doi: 10.1021/acssynbio.9b00523.

[93] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, M. J. Schilstra, P. J. C. Clarke, A. G. Rust, Z. Pan, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology*, 246(1):162–190, June 2002. ISSN 0012-1606.

doi: 10.1006/dbio.2002.0635. URL `https://www.sciencedirect.com/science/article/pii/S0012160602906354`.

[94] D. C. Lyons, S. L. Kaltenbach, and D. R. McClay. Morphogenesis in sea urchin embryos: Linking cellular events to gene regulatory network states. *Wiley Interdisciplinary Reviews: Developmental Biology*, 1(2):231–252, April 2012. ISSN 1759-7692. doi: 10.1002/wdev.18.

[95] A. Streit, M. Tambalo, J. Chen, T. Grocott, M. Anwar, A. Sosinsky, and C. D. Stern. Experimental approaches for gene regulatory network construction: the chick as a model system. *Genesis*, 51(5):296–310, May 2013. ISSN 1526-968X. doi: 10.1002/dvg.22359.

[96] R. L. Levine, M. Wadleigh, J. Cools, B. L. Ebert, G. Wernig, B. J. P. Huntly, T. J. Boggon, L. Wlodarska, J. J. Clark, S. Moore, J. Adelsperger, S. Koo, J. C. Lee, S. Gabriel, T. Mercher, A. D'Andrea, S. Frohling, K. Dohner, P. Marynen, P. Vandenberghe, R. A. Mesa, A. Tefferi, J. D. Griffin, M. J. Eck, W. R. Sellers, M. Meyerson, T. R. Golub, S. J. Lee, and D. G. Gilliland. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell*, 7(4):387–397, 2005. ISSN 1535-6108. doi: 10.1016/J.Ccr.2005.03.023.

[97] T. I. Lee and R. A. Young. Transcriptional regulation and its misregulation in disease. *Cell*, 152(6):1237–1251, March 2013. ISSN 0092-8674. doi: 10.1016/j.cell.2013.02.014. URL `https://www.sciencedirect.com/science/article/pii/S0092867413002031`.

[98] K. M. Cherry and L. Qian. Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks. *Nature*, 559(7714):370–376, July 2018. ISSN 1476-4687. doi: 10.1038/s41586-018-0289-6. URL `https://www.nature.com/articles/s41586-018-0289-6`.

[99] L. Qian, E. Winfree, and J. Bruck. Neural network computation with DNA strand displacement cascades. *Nature*, 475(7356):368–372, 2011. ISSN 0028-0836. doi: 10.1038/Nature10262.

[100] J. Kuo, R. Yuan, C. Sánchez, J. Paulsson, and P. A. Silver. Toward a translationally independent RNA-based synthetic oscillator using deactivated CRISPR-Cas. *Nucleic Acids Research*, 48(14):8165–8177, August 2020. ISSN 0305-1048. doi: 10.1093/nar/gkaa557. URL `https://doi.org/10.1093/nar/gkaa557`.

[101] B. H. Weinberg, N. T. H. Pham, L. D. Caraballo, T. Lozanoski, A. Engel, S. Bhatia, and W. W. Wong. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nature Biotechnology*, 35(5):453–462, May 2017. ISSN 1546-1696. doi: 10.1038/nbt.3805. URL `https://www.nature.com/articles/nbt.3805`.

[102] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu. Synthetic recombinase-based state machines in living cells. *Science*, 353 (6297):aad8559, July 2016. doi: 10.1126/science.aad8559. URL `https://www.science.org/doi/10.1126/science.aad8559`.

[103] S. Ausländer, D. Ausländer, M. Müller, M. Wieland, and M. Fussenegger. Programmable single-cell mammalian biocomputers. *Nature*, 487(7405): 123–127, July 2012. ISSN 1476-4687. doi: 10.1038/nature11149. URL `https://www.nature.com/articles/nature11149`.

[104] L. M. Hochrein, H. Li, and N. A. Pierce. High-performance allosteric conditional guide RNAs for mammalian cell-selective regulation of CRISPR/Cas. *ACS Synthetic Biology*, 10(5):964–971, 2021.

[105] E. H.-Y. Yeung. *Reverse Engineering and Quantifying Context Effects in Synthetic Gene Networks*. PhD, California Institute of Technology, 2016. URL `https://resolver.caltech.edu/CaltechThesis:05272016-145559554`.

[106] J. C. Anderson. Anderson promoter collection, 2006. URL `http://parts.igem.org/Promoters/Catalog/Anderson`.

[107] S. Tiplady. Lateral Flow and Consumer Diagnostics. In *The Immunoassay Handbook*, pages 533–536. Elsevier, 2013. ISBN 978-0-08-097037-0.

[108] J. Larsen, P. Buchanan, S. Johnson, S. Godbert, and M. Zinaman. Human chorionic gonadotropin as a measure of pregnancy duration. *International Journal of Gynecology & Obstetrics*, 123(3):189–195, 2013. ISSN 1879-3479. doi: 10.1016/j.ijgo.2013.05.028.

[109] C. Gnoth and S. Johnson. Strips of hope: Accuracy of home pregnancy tests and new developments. *Geburtshilfe Frauenheilkd*, 74(07):661–669, August 2014. ISSN 0016-5751, 1438-8804. doi: 10.1055/s-0034-1368589.

[110] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo. SARS-CoV-2 (COVID-19) by the numbers. *eLife*, 9:e57309, March 2020. ISSN 2050-084X. doi: 10.7554/eLife.57309.

[111] K. K.-W. To, O. T.-Y. Tsang, C. C.-Y. Yip, K.-H. Chan, T.-C. Wu, J. M.-C. Chan, W.-S. Leung, T. S.-H. Chik, C. Y.-C. Choi, D. H. Kandamby, D. C. Lung, A. R. Tam, R. W.-S. Poon, A. Y.-.F. Fung, I. F.-N. Hung, V. C.-C. Cheng, J. F.-W. Chan, and K.-Y. Yuen. Consistent Detection of 2019 Novel Coronavirus in Saliva. *Clinical Infectious Diseases*, 71(15):841–843, February 2020. ISSN 1058-4838, 1537-6591. doi: 10.1093/cid/ciaa149.

[112] K. K.-W. To, O. T.-Y. Tsang, W.-S. Leung, A. R. Tam, T.-C. Wu, D. C. Lung, C. C.-Y. Yip, J.-P. Cai, J. M.-C. Chan, T. S.-H. Chik, D. P.-L. Lau, C. Y.-C. Choi, L.-L. Chen, W.-M. Chan, K.-H. Chan, J. D. Ip, A. C.-K.

Ng, R. W.-S. Poon, C.-T. Luo, V. C.-C. Cheng, J. F.-W. Chan, I. F.-N. Hung, Z. Chen, H. Chen, and K.-Y. Yuen. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: An observational cohort study. *Lancet Infectious Diseases*, 20:565–574, March 2020. ISSN 14733099. doi: 10.1016/S1473-3099(20) 30196-1.

[113] V. M. Corman, O. Landt, M. Kaiser, R. Molenkamp, A. Meijer, D. K. W. Chu, T. Bleicker, S. Brünink, J. Schneider, M. L. Schmidt, D. G. J. C. Mulders, B. L. Haagmans, B. van der Veer, S. van den Brink, L. Wijsman, G. Goderski, J.-L. Romette, J. Ellis, M. Zambon, M. Peiris, H. Goossens, C. Reusken, M. P. G. Koopmans, and C. Drosten. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance*, 25(3), January 2020. ISSN 1560-7917. doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

[114] J. F.-W. Chan, C. C.-Y. Yip, K. K.-W. To, T. H.-C. Tang, S. C.-Y. Wong, K.-H. Leung, A. Y.-F. Fung, A. C.-K. Ng, Z. Zou, H.-W. Tsoi, G. K.-Y. Choi, A. R. Tam, V. C.-C. Cheng, K.-H. Chan, O. T.-Y. Tsang, and K.-Y. Yuen. Improved molecular diagnosis of COVID-19 by the novel, highly sensitive and specific COVID-19-RdRp/Hel real-time reverse transcription-polymerase chain reaction assay validated *in vitro* and with clinical specimens. *Journal of Clinical Microbiology*, 58(5):1–10, 2020. ISSN 0095-1137, 1098-660X. doi: 10.1128/JCM.00310-20.

[115] E. Fu, T. Liang, P. Spicar-Mihalic, J. Houghtaling, S. Ramachandran, and P. Yager. Two-dimensional paper network format that enables simple multi-step assays for use in low-resource settings in the context of malaria antigen detection. *Analytical Chemistry*, 84(10):4574–4579, May 2012. ISSN 0003-2700, 1520-6882. doi: 10.1021/ac300689s.

[116] S. Ramachandran, E. Fu, B. Lutz, and P. Yager. Long-term dry storage of an enzyme-based reagent system for ELISA in point-of-care devices. *Analyst*, 139(6):1456–1462, 2014. ISSN 0003-2654, 1364-5528. doi: 10.1039/C3AN02296J.

[117] S. J. Schulte, J. Huang, and N. A. Pierce. HCR lateral flow assays for amplified instrument-free at-home SARS-CoV-2 testing, September 2022. URL https://www.biorxiv.org/content/10.1101/2022.09.18.508442v1.

[118] D. Raoult, A. Zumla, F. Locatelli, G. Ippolito, and G. Kroemer. Coronavirus infections: Epidemiological, clinical and immunological features and hypotheses. *Cell Stress*, 4(4):66–75, April 2020. ISSN 25230204. doi: 10.15698/cst2020.04.216.

[119] D. Wang, S. He, X. Wang, Y. Yan, J. Liu, S. Wu, S. Liu, Y. Lei, M. Chen, L. Li, J. Zhang, L. Zhang, X. Hu, X. Zheng, J. Bai, Y. Zhang, Y. Zhang, M. Song,

and Y. Tang. Rapid lateral flow immunoassay for the fluorescence detection of SARS-CoV-2 RNA. *Nature Biomedical Engineering*, 4(12):1150–1158, December 2020. ISSN 2157-846X. doi: 10.1038/s41551-020-00655-z.

[120] L. C. Tindale, J. E. Stockdale, M. Coombe, E. S. Garlock, W. Y. V. Lau, M. Saraswat, L. Zhang, D. Chen, J. Wallinga, and C. Colijn. Evidence for transmission of COVID-19 prior to symptom onset. *eLife*, 9:e57149, June 2020. ISSN 2050-084X. doi: 10.7554/eLife.57149.

[121] M. A. Johansson, T. M. Quandelacy, S. Kada, P. V. Prasad, M. Steele, J. T. Brooks, R. B. Slayton, M. Biggerstaff, and J. C. Butler. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Network Open*, 4(1):e2035057, January 2021. ISSN 2574-3805. doi: 10.1001/jamanetworkopen.2020.35057.

[122] J. N. Zadeh, C. D. Steenberg, J. S. Bois, B. R. Wolfe, M. B. Pierce, A. R. Khan, R. M. Dirks, and N. A. Pierce. NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, 32(1):170–173, 2011. doi: 10.1002/jcc.21596.

[123] R. M. Dirks and N. A. Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *Journal of Computational Chemistry*, 24:1664–1677, 2003.

[124] R. M. Dirks, M. Lin, E. Winfree, and N. A. Pierce. Paradigms for computational nucleic acid design. *Nucleic Acids Research*, 32(4):1392–1403, 2004.

[125] R. M. Dirks and N. A. Pierce. An algorithm for computing nucleic acid base-pairing probabilities including pseudoknots. *Journal of Computational Chemistry*, 25:1295–1304, 2004.

[126] J. N. Zadeh, B. R. Wolfe, and N. A. Pierce. Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of Computational Chemistry*, 32:439–452, 2011. doi: 10.1002/jcc.21633.

[127] I. Tinoco, Jr., O. C. Uhlenbeck, and M. D. Levine. Estimation of secondary structure in ribonucleic acids. *Nature*, 230:362–367, 1971.

[128] J. SantaLucia, Jr. A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics. *Proceedings of the National academy of Sciences*, 95(4):1460–1465, 1998.

[129] T. B. Xia, J. SantaLucia, Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Q. Jiao, C. Cox, and D. H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37(42):14719–14735, 1998.

[130] S. Bommarito, N. Peyret, and J. SantaLucia. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Research*, 28(9): 1929–1934, 2000. ISSN 0305-1048. doi: 10.1093/nar/28.9.1929.

[131] N. Peyret. *Prediction of Nucleic Acid Hybridization: Parameters and Algorithms*. PhD, Wayne State University, 2000.

[132] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, 31(13):3406–3415, 2003.

[133] J. SantaLucia, Jr. and D. Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440, 2004. ISSN 1056-8700.

[134] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, and D. H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National academy of Sciences*, 101(19):7287–7292, 2004. ISSN 0027-8424. doi: 10.1073/pnas.0401799101.

[135] R. T. Koehler and N. Peyret. Thermodynamic properties of dna sequences: Characteristic values for the human genome. *Bioinformatics*, 21(16):3333–3339, 2005. ISSN 1367-4803. doi: 10.1093/Bioinformatics/Bti530.

[136] Z. J. Lu, D. H. Turner, and D. H. Mathews. A set of nearest neighbor parameters for predicting the enthalpy change of RNA secondary structure formation. *Nucleic Acids Research*, 34(17):4912–4924, 2006. ISSN 0305-1048. doi: 10.1093/nar/gkl472.

[137] D. H. Turner and D. H. Mathews. NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*, 38:D280–D282, 2010. ISSN 0305-1048. doi: 10.1093/nar/gkp892.

[138] C. Geary, P. W. K. Rothemund, and E. S. Andersen. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science*, 345(6198):799–804, 2014.

[139] A. D. Halleran, A. Swaminathan, and R. M. Murray. Single day construction of multigene circuits with 3G assembly. *ACS Synthetic Biology*, 7(5):1477–1480, May 2018. doi: 10.1021/acssynbio.8b00060. URL https://doi.org/10.1021/acssynbio.8b00060.

[140] S. V. Iverson, T. L. Haddock, J. Beal, and D. M. Densmore. CIDAR MoClo: Improved MoClo assembly standard and new E. coli part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synthetic Biology*, 5(1):99–103, January 2016. doi: 10.1021/acssynbio.5b00124. URL https://doi.org/10.1021/acssynbio.5b00124.

[141] F. St-Pierre, L. Cui, D. G. Priest, D. Endy, I. B. Dodd, and K. E. Shear-win. One-step cloning and chromosomal integration of DNA. *ACS Synthetic Biology*, 2(9):537–541, September 2013. doi: 10.1021/sb400021j. URL `https://doi.org/10.1021/sb400021j`.

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

A.1 Materials and methods

Plasmid construction and molecular cloning

Sequences for parts used are provided in Section A.2. Plasmid layouts for each construct as well as representative plasmid maps and corresponding full plasmid sequences are provided in Section A.3. For plasmid construction, all extension and inverse PCR were performed with Q5 Hot Start High-Fidelity 2X Master Mix (NEB, M0494S). Transformation and selection during cloning were performed in either NEB® 10-beta Competent *E. coli* (NEB, C3019H) or NEB 5-alpha Competent *E. coli* (NEB, C2987H). Plasmid sequences were confirmed via Sanger sequencing with Laragen Corp. Plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen, 27106X4). Gel extractions are performed with MinElute Gel Extraction Kit (Qiagen, 28604). Gibson assembly was performed with NEBuilder® HiFi DNA Assembly Master Mix (NEB, E2621S).

Circuit constructs were generated via Gibson assembly of transcriptional units. Gibson overhangs from the 3G cloning technique were used [139]. For all circuit plasmids, transcriptional units were assembled on either a pDVA vector or pDVS vector. pDVA_colE1 vector was generated by extension PCR of a MoClo vector (gift from D. Densmore, Addgene plasmid #66040) [140] by adding Gibson overlap regions. pDVS_psc vector was generated by modification of a pDVA_psc vector (gift from R. Murray) to replace the carbenicillin resistance gene with an *aadA* gene cloned from a JEN202 cell strain (gift from L. Marraffini) [67].

Transcriptional units consists of promoter, cgRNA or trigger, and terminators. Parts plasmids containing transcriptional units were generated by performing inverse PCR on cgRNA parts plasmids from a previous study [1]. The linear transcriptional units were generated using extension PCR to append appropriate Gibson overhangs. Linearized vectors were generated using extension-inverse PCR to append appropriate Gibson overhangs to pDVA_colE1 or pDVS_psc. Fragments were purified by gel extraction and assembled via Gibson assembly.

Genomic Incorporation of mRFP1 into Marionette strain

A transcriptional unit consisting of constitutively expressed mRFP1 is genomically incorporated into WT Marionette cells (gift from C. Voigt) [61] using pOSIP [141]. Briefly, the insert fragment was obtained by extension PCR performed to extract an mRFP1 transcriptional unit with Gibson overhang extensions from the previously obtained Ec001 cell lines (gift from S. Qi) [70]. The vector fragment was obtained by digesting the pOSIP-KO plasmid (gift from K. E. Shearwin, Addgene plasmid #45985) [141] with *SpeI* and *BamHI*. Fragments are gel purified and combined via Gibson assembly. Final plasmid product was transformed into WT-Marionette cells via chemical transformation and plated overnight on Agar plates with Kanamycin.

Cell culture

Circuit plasmids were transformed one or two at a time using chemical transformation. Briefly, the cell strain was grown overnight in LB, then sub-cultured with a 20 to 100-fold dilution for 3-5 hours. The sub-culture was centrifuged, and re-suspended in TSS buffer (Thermo Scientific, J62318EQE). Cells were transformed immediately or stored at -80 °C. All cell culturing steps were performed at 37°C. Antibiotic working concentrations used were 100μg/mL, 30μg/mL, 34μg/mL, and 50μg/mL for carbenicillin, kanamycin, chloramphenicol, and spectinomycin, respectively.

Induction assay for thresholding switch

Individual colonies from circuit transformations were picked and grown in 1mL EZ-RDM buffer with appropriate antibiotics overnight. Cultures were further diluted 100-fold in fresh EZ-RDM with 20nM anhydrotetracycline (aTc, Takara, 631310) and cultured for 4 hours. OD600 values were measured for subcultures and all cultures were adjusted to the cell strain with the lowest concentration. Cells are diluted again 100-fold into EZ-RDM with 20nM aTc with 5000, 1250, 312, 78, 20, and 0μM IPTG. Cell growth and fluorescence kinetics were read with a microplate reader (Biotek Synergy Neo2).

Induction assay for toggle switch

Individual colonies from circuit transformations were picked and grown in 1mL EZ-RDM buffer with appropriate antibiotics overnight. Cultures were further diluted 100-fold in fresh EZ-RDM with 20nM aTc and cultured for 4 hours. Cells were cultured in a microplate reader for 12 hours in EZ-RDM with appropriate antibiotics and 20nM aTc containing all cgRNA and trigger inducers. At the end of the

induction, 200μL of cells were filtered through a 30-40 μm cell filter (Pall, 8027) and its fluorescence measured using flow cytometry. Further, cells were diluted again in 100-fold in fresh EZ-RDM with appropriate antibiotics and 20nM aTc containing cgRNA inducers, but no trigger inducers. At the end of induction, cells were filtered and measured using flow cytometry. The above two steps were then repeated starting with the subculture at the end of the 2nd incubation period, with swapped trigger inducer conditions to demonstrate switching behavior.

Flow cytometry

Protein fluorescence was measured with the CytoFLEX S Flow Cytometer (Beckman Coulter) in the Caltech Flow Cytometry Facility. Over 100,000 live cells were collected for each well at a flow rate of 10 μL/min. mRFP1 fluorescence was measured using the 561 nm laser with the 585/42 nm filter. SSC-H vs. FSC-H was used to gate for live cells (see example in Figure A.1). A value of 700 was added to all fluorescence values in the 585Yellow-A channel to remove most negative fluorescence values. The first 10,000 data points from the gated area were plotted in a logarithmic histogram to generate flow cytometry figures.
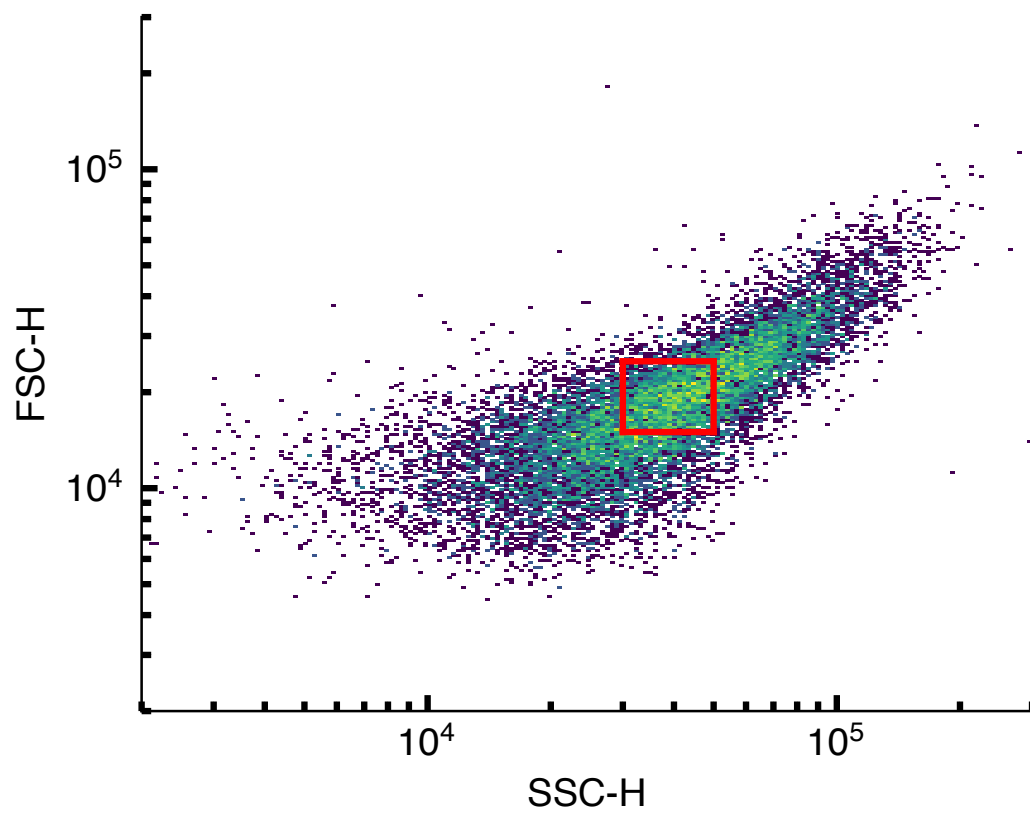
Figure A.1: Illustration of the gates used for flow cytometry analysis of cells. Example of gating for cells on SSC-H vs. FSC-H.

A.2 Sequences for cgRNAs, triggers, and plasmid elements

	Name	Sequence	Legend
a	**Splinted switch cgRNAs**		
	cgRNA D	5′-CATCTAATTCAACAAGAATTGTTTTAGAGCTA CACCTTACGCCGGTTCAATTCCAAGTCCCTTCCAG TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT AACACCCTTTACAAACCTTCCTCTTCCTTTACCCT AAGTGGCACCGAGTCGGTGCTTTTTT-3′	GFP targeting cgRNA D
	cgRNA B	5′-AACTTTCAGTTTAGCGGTCTGTTTTAGAGCTA GTCGTTACCTTATCAATATCAACCTCCGCATACAC TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GCACATAGGACCCAACATGCCAACAGAGAAGAGTT AAGTGGCACCGAGTCGGTGCTTTTTT-3′	RFP targeting cgRNA B
b	**Target sites for cgRNA**		
	GFP Target + PAM	5′-CCCAATTCTTGTTGAATTAGATG-3′	GFP target and PAM site fused to the 5′ end of some triggers or cgRNAs
	RFP Target + PAM	5′-CCCAGACCGCTAAACTGAAAGTT-3′	RFP target and PAM site fused to the 5′ end of some triggers or cgRNAs
c	**Splinted switch triggers**		
	Trigger D	5′-AGGGTAAAGGAAGAGGAAGGTTTGTAAAGGGTGTT CTGGAAGGGACTTGGAATTGAACCGGCGTAAGGTG-3′	Trigger D
	Trigger B	5′-AACTCTTCTCTGTTGGCATGTTGGGTCCTATGTGC GTGTATGCGGAGGTTGATATTGATAAGGTAACGAC-3′	Trigger B
	Trigger C	5′-GACTGTATAAGAATGAACTACTTCAGACCCGTATG TTATGACGATGGGAAATTTACTATGATTCGATTAC-3′	Trigger C

Table A.1: Control gRNA and trigger sequences. (a) Orthogonal cgRNA sequences. (b) 5′ target sites for cgRNA transcriptional regulation. (c) Orthogonal trigger sequences.

Name	Sequence	Legend
Promoters and terminators		
B1006 Terminator	5′-aaaaaaaaaccccgcccctgacagggcgggg ttttttttt-3′	Short terminator for cgRNAs and triggers
pLac	5′-AATTGTGAGCGGATAACAATTGACATTGT GAGCGGATAACAAGATACTGAGCACA-3′	pLac promoter R0011
J23105 Const. Promoter	5′-tttacggctagctcagtcctaggtacta tgctagc-3′	J23105 promoter [106]
J23106 Const. Promoter	5′-tttacggctagctcagtcctaggtatag tgctagc-3′	J23106 promoter [106]
pTac	5′-TGTTGACAATTAATCATCGGCTCGTAT AATGTGTGGAATTGTGAGCGCTCACAATT-3′	pTac promoter
pSalTTC	5′-GGGGCCTCGCTTGGGTTATTGCTGGTGCCCGGCC GGGCGCAATATTCATGTTGATGATTTATTATATATCG AGTGGTGTATTTATTTATATTGTTTGCTCCG TTACCGTTATTAAC-3′	pSalTTC promoter
pLuxB	5′-ACCTGTAGGATCGTACAGGTTTACGCAA GAAAATGGTTTGTTACAGTCGAATAAA-3′	pLuxB promoter
pCin	5′-CCCTTTGTGCGTCCAAACGGACGCACGGCGCTCT AAAGCGGGTCGCGATCTTTCAGATTCGCTCCTCGCGC TTTCAGTCTTTGTTTTGGCGCATGTCGTTATCGCAAAA CCGCTGCACACTTTTGCGCGACATGCTCTGATCCCCCT CATCTGGGGGGGCCTATCTGAGGGAATTTCCGATCCGG CTCGCCTGAACCATTCTGCTTTCCACGAACTTGAAAAC GCT-3′	pCin promoter
J23110 Const. Promoter	5′-tttacggctagctcagtcctaggtacaa tgctagc-3′	J23110 promoter [106]
J23116 Const. Promoter	5′-ttgacagctagctcagtcctagggacta tgctagc-3′	J23116 promoter [106]
J23117 Const. Promoter	5′-ttgacagctagctcagtcctagggattg tgctagc-3′	J23117 promoter [106]

Table A.2: Plasmid element sequences. Promoter and terminator element sequences.

A.3   Representative plasmid maps and annotated sequences

Table A.3 describes the plasmid constructs used for in this study. Representative plasmid maps and annotated sequences are provided in Figures A.2–A.11. Plasmids for additional studies were constructed analogously.

	Name	Parts	Figures	Legend
a	**Plasmids used in threshold switch studies**			
	pdCas9-LacI	Previous study [1]	2.5, 2.7, 2.6, 2.9	
	P1	pSC101; spec-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger B–B1006; J23105–Trigger C–B1006;	2.7, 2.9	Closed loop
	P2	pSC101; spec-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger C–B1006; J23105–Trigger C–B1006;	2.7, 2.9	Open loop by hybridization
	P3	pSC101; spec-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger C–B1006; J23105–Trigger B–B1006;	2.7, 2.9	Open loop by transcription
	P4	ColE1; amp-R; pLac–GFP targeting cgRNA D–B1006; J23105–GFP Target + PAM–Trigger D–B1006; J23105–Trigger C–B1006;	2.5, 2.9	Closed loop
	P5	ColE1; amp-R; pLac–GFP targeting cgRNA D–B1006; J23105–GFP Target + PAM–Trigger C–B1006; J23105–Trigger C–B1006;	2.5, 2.9	Open loop by hybridization
	P6	ColE1; amp-R; pLac–GFP targeting cgRNA D–B1006; J23105–GFP Target + PAM–Trigger C–B1006; J23105–Trigger D–B1006;	2.5, 2.9	Open loop by transcription
b	**Plasmids used in toggle switch studies**			
	pdCas9-bacteria	Addgene plasmid #44251	2.12	
	P7	ColE1; amp-R; pLuxB–GFP Target + PAM–RFP targeting cgRNA B–B1006; pCin–Trigger B–B1006	2.12	RFP half of toggle switch
	P8	pSC101; spec-R; pTac–RFP Target + PAM–GFP targeting cgRNA D–B1006; pSalTTC–Trigger D–B1006	2.12	GFP half of toggle switch

Table A.3: Plasmids used for cgRNA circuits in *E. coli*.

Name	Parts	Legend
Plasmids used in threshold switch optimization studies		
P9	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23110–RFP Target + PAM–Trigger B–B1006; J23110–GFP Target + PAM–Trigger C–B1006;	J23110 promoter, closed loop
P10	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23110–RFP Target + PAM–Trigger C–B1006; J23110–GFP Target + PAM–Trigger C–B1006;	J23110 promoter, open loop by hybridization
P11	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23110–RFP Target + PAM–Trigger C–B1006; J23110–GFP Target + PAM–Trigger B–B1006;	J23110 promoter, open loop by transcription
P12	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger B–B1006; J23105–GFP Target + PAM–Trigger C–B1006;	J23105 promoter, closed loop
P13	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger C–B1006; J23105–GFP Target + PAM–Trigger C–B1006;	J23105 promoter, open loop by hybridization
P14	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23105–RFP Target + PAM–Trigger C–B1006; J23105–GFP Target + PAM–Trigger B–B1006;	J23105 promoter, open loop by transcription
P15	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23116–RFP Target + PAM–Trigger B–B1006; J23116–GFP Target + PAM–Trigger C–B1006;	J23116 promoter, closed loop
P16	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23116–RFP Target + PAM–Trigger C–B1006; J23116–GFP Target + PAM–Trigger C–B1006;	J23116 promoter, open loop by hybridization
P17	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23116–RFP Target + PAM–Trigger C–B1006; J23116–GFP Target + PAM–Trigger B–B1006;	J23116 promoter, open loop by transcription
P18	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23117–RFP Target + PAM–Trigger B–B1006; J23117–GFP Target + PAM–Trigger C–B1006;	J23117 promoter, closed loop
P19	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23117–RFP Target + PAM–Trigger C–B1006; J23117–GFP Target + PAM–Trigger C–B1006;	J23117 promoter, open loop by hybridization
P20	ColE1; amp-R; pLac–RFP targeting cgRNA B–B1006; J23117–RFP Target + PAM–Trigger C–B1006; J23117–GFP Target + PAM–Trigger B–B1006;	J23117 promoter, open loop by transcription

Table A.4: Plasmids used for threshold switch optimization studies in *E. coli*.

Figure A.2: Plasmid map for P1.

••• TAAGCGAGGCCGCCCGACTGATACGTTGATTTTCCAAGTTGAACTAGATAGACAAATGGATCTCG 65
TAACCGAACTTGAGAACAACCAGATAAAAATGAATGGTGACAAAATACCAACAACCATTACATCA 130
GATTCCTACCTACGTAACGGACTAAGAAAAACACTACACGATGCTTTAACTGCAAAAATTCAGCT 195
CACCAGTTTTGAGGCAAAATTTTTGAGTGACATGCAAAGTAAGCATGATCTCAATGGTTCGTTCT 260
CATGGCTCACGCAAAAACAACGAACCACACTAGAGAACATACTGGCTAAATACGGAAGGATCTGA 325
GGTTCTTATGGCTCTTGTATCTATCAGTGAAGCATCAAGACTAACAAACAAAAGTAGAACAACTG 390
TTCACCGTTAGATATCAAAGGGAAAACTGTCCATATGCACAGATGAAAACGGTGTAAAAAAGATA 455
GATACATCAGAGCTTTTACGAGTTTTTGGTGCATTTAAAGCTGTTCACCATGAACAGATCGACAA 520
TGTAACtactagagctaaatacattcaaatatctatccgctcatgagacaaataaccctgataaat 585
gcttcaataatattgaaaaaggaagaatATGAGTGAAAAAGTGCCCGCCGAGATTTCGGTGCAAC 650
TATCACAAGCACTCAACGTCATCGGGCGCCACTTGGAGTCGACGTTGCTGGCCGTGCATTTGTAC 715
GGCTCCGCACTGGATGGCGGATTGAAACCGTACAGTGATATTGATTTGCTGGTGACTGTAGCTGC 780
ACCGCTCAATGATGCCGTGCGGCAAGCCCTGCTCGTCGATCTCTTGGAGGTTTCAGCTTCCCCTG 845
GCCAAAACAAGGCACTCCGCGCCTTGGAAGTGACCATCGTCGTGCACAGTGACATCGTACCTTGG 910
CGTTATCCGGCCAGGCGGGAACTGCAGTTCGGAGAGTGGCAGCGCAAAAGACATCCTTGCGGGCAT 975
CTTCGAGCCCGCCACAACCGATTCTGACTTGGCGATTCTGCTAACAAAGGCAAAGCAACATAGCG 1040
TCGTCTTGGCAGGTTCAGCAGCGAAGGATCTCTTCAGCTCAGTCCCAGAAAGCGATCTATTCAAG 1105
GCACTGGCCGATACTCTGAAGCTATGGAACTCGCCGCCAGATTGGGCGGGGGATGAGCGGAATGT 1170
AGTGCTTACTTTGTCTCGTATCTGGTACACCGCAGCAACCGGCAAGATCGCGCCAAAGGATGTTG 1235
CTGCCACTTGGGCAATGGCACGCTTGCCAGCTCAACATCAGCCCATCCTGTTGAATGCCAAGCGG 1300
GCTTATCTTGGGCAAGAAGAAGATTATTTGCCCGCTCGTGCGGATCAGGTGGCGGCGCTCATTAA 1365
ATTCGTGAAGTATGAAGCAGTTAAACTGCTTGGTGCCAGCCAATGAtaatactagctccggcaaa 1430
aaaacgggcaaggtgtcaccaccctgcccttttctctttaaaaccgaaaagattacttcgcgaaat 1495
aggcgtatcacgaggcagaatttcagataaaaaaaatcctttagctttcgctaaggatgatttctg 1560
gaattcgcggccgcttctagagactagtggAAGACACATTACTCGCATCCATTCTCAGGCTGTCT 1625
CGTCTCGTCTCgAATTCGTGGTAGGAGATCTGAAACATGACGAGCTGGTGAGCATAAGGAGAATT 1690
GTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACAAACTTTCAGTTTAG 1755
CGGTCTGTTTTAGAGCTAGTCGTTACCTTATCAATATCAACCTCCGCATACACTAGCAAGTTAAA 1820
ATAAGGCTAGTCCGTTATCAACTTGCACATAGGACCCAACATGCCAACAGAGAAGAGTTAAGTGG 1885
CACCGAGTCGGTGCTTTTTTTTaaaaaaaaaacccccgcccctgacagggcggggttttttttGCGCT 1950
ATCAGTTGACACAGTGAAGCTTGGGCCCGAACAAAAACTCatctcagaagaggatctgaatagcg 2015
ccgtcgaccatcatcatcatcatcattgagtttaaacggtctccagcttggctgttttggcggat 2080
gagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctgataaaaacagaat 2145
ttgcctggcggcagtagcgcggtggtcccacctgacccccatgccgaactcagaagtgaaacgccg 2210
tagcgccgatggtagtgtgggggtctcccatgcgagagtagggaactgccaggcatcaaataaaa 2275
cgaaaggctcagtcgaaagactGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACTGGATCCT 2340
TACTCGAGTCTAGGCACTGAAGGTCCTCAATCGCACTGGAAACATCAAGGTCGGAATTCGTGGTA 2405
GGAGATCTGCATAAGGAGtttacggctagctcagtcctaggtactatgctagccccAGACCGCTA 2470
AACTGAAAGTTAACTCTTCTCTGTTGGCATGTTGGGTCCTATGTGCGTGTATGCGGAGGTTGATA 2535
TTGATAAGGTAACGACaaaaaaaaaccccgcccctgacagggcggggtttttttttCTACACCCTT 2600
CTAACAACTCcccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttat 2665
ctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttct 2730
gcgttttataGGATCCTTACTCGAGTCTAGCTGACCTCCTGCCAGCAATAGTAAGACAACACGCAA 2795
AGTCGAATTCGTGGTAGGAGATCTGCATAAGGAGttacggctagctcagtcctaggtactatgc 2860
tagcGACTGTATAAGAATGAACTACTTCAGACCCGTATGTTATGACGATGGGAAATTTACTATGA 2925
TTCGATTACaaaaaaaaaacccgcccctgacagggcggggtttttttttCTACACCCTTCTAACAA 2990
CTCcccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgt 3055
ttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgttta 3120
taGGATCCTTACTCGAGTCTAGCCAGGATACATAGATTACCACAACTCCGAGCCCTTCCACCTGT 3185
CTTCtactagtagcggccgctgcagtccggcaaaaaagggcaaggtgtcaccaccctgcccttttt 3250
tctttaaaaccgaaaagattacttcgcgttcctcgctcactgactcgctgcgctc 3315
ggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaattctgctagagCTGTCAG 3380
ACCAAGTTTACGGCTCGCTTGGACTCCTGTTGATAAGGATCCAGTAATGACCTCAGAACTCCATCT 3445
GGATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCACTAGGG 3510
ACAGTAAGACGGGTAAGCCTGTTGATGATACCGCTGCCTTACTGGGTGCATTAGCCAGTCTGAAT 3575
GACCTGTCACGGGATAATCCGAAGTGGTCAGACTGGAAAATCAGAGGGCAGGAACTGCTGAACAG 3640
CAAAAAGTCAGATAGCACCACATAGCAGACCCGCACCATGGAAGCCCCTGAGAAGCCCGTGACGGGC 3705
TTTTCTTGTATTATGGGTAGTTTCCTTGCATGAATCCATAAAAGGCGCCTGTAGTGCCATTTACC 3770
CCCATTCACTGCCAGACCGTGAGCGGCAGCGAACTGAATGTCACGAAAAAGACAGCGACTCAGGT 3835
GCCTGATGGTCGGAGACAAAAGGAATATTCAGCGATTTGCCCGAGCTTGCGAGGGTGCTACTTAA 3900
GCCTTTAGGGTTTTAAGGTCTGTTTTGTAGAGGAGCAAACAGCGTTTGCGACATCCTTTTGTAAT 3965
ACTGCGGAACTGACTAAAGTAGTGAGTTATACACAGGGCTGGGATCTATTCTTTTTATCTTTTTT 4030
TATTCTTTCTTTATTCTATAAATTATAACCACTTGAATATAAACAAAAAAAACACACAAAGGTCT 4095
AGCGGAATTTACAGAGGGTCTAGCAGAATTTACAAGTTTTCCAGCAAAGGTCTAGCAGAATTTAC 4160
AGATACCCACAACTCAAAGGAAAAGGACATGTAATTATCATTGACTAGCCCATCTCAATTGGTAT 4225
AGTGATTAAAATCACCTAGACCAATTGAGATGTATGTCTGAATTAGTTGTTTTCAAAGCAAATTA 4290
ACTAGCGATTAGTCGCTATGACTTAACGGAGCATGAAACCAAGCTAATTTTATGCTGTGTGGCAC 4355
TACTCAACCCCACGATTGAAAACCCTACAAGGAAAGAGGACGGTATCGTTCACTTATAACCAA 4420
TACGCTCAGATGATGAACATCAGTAGGGAAAATGCTTATGGTGTATTAGCTAAAGCAACCAGAGA 4485
GCTGATGACGAGAACTGTGGAAATCAGGAATCCTTTGGTTAAAGGCTTTGAGATTTTCCAGTGGA 4550
CAAACTATGCCAAGTTCTCAAGCGAAAAATTAGAATTAGTTTTTAGTGAAGAGATATTGCCTTAT 4615
CTTTTCCAGTTAAAAAAATTCATAAAAATATAATCTGGAACATGTTAAGTCTTTTGAAAACAAATA 4680
CTCTATAGGAGATTTATGAGTGGTTATTAAAAGAACTAACACAAAAGAAAACTCACAAGGCAAATA 4745
TAGAGATTAGCCTTGATGAATTTAAGTTCATGTTAATGCTTGAAAATAACTACCATGAGTTTAAA 4810
AGGCTTAACCAATGGGTTTTGAAACCAATAAGTAAAGATTTAAACACTTACAGCAATATGAAATT 4875
GGTGGTTGA ••• 4884
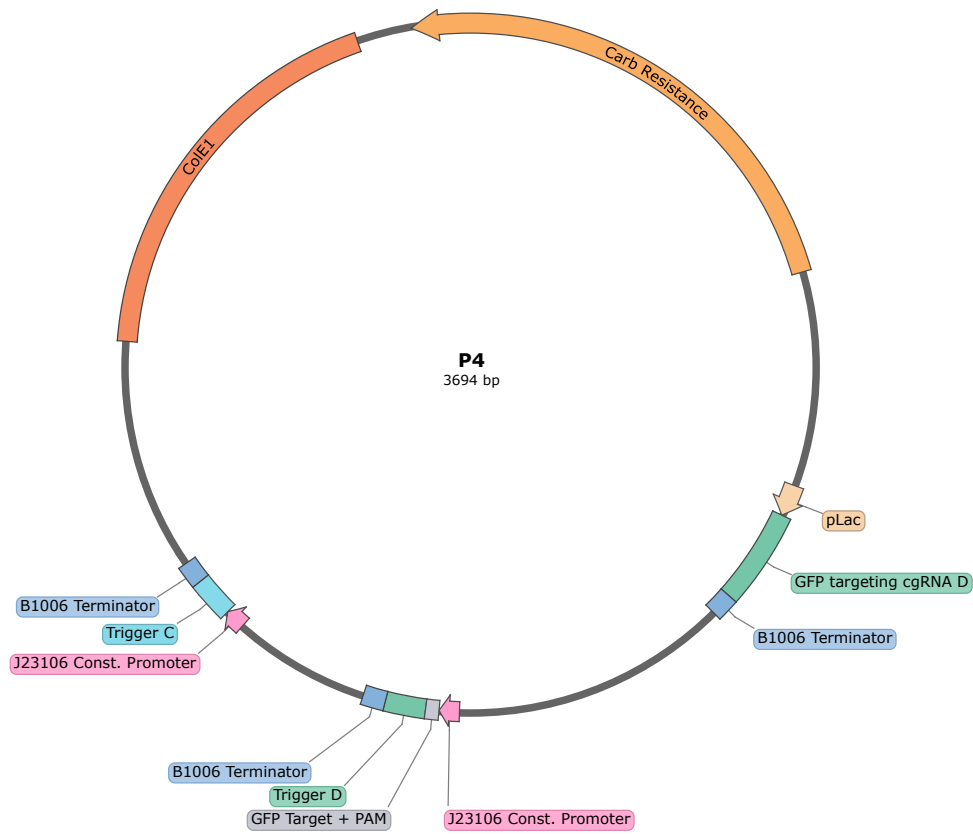
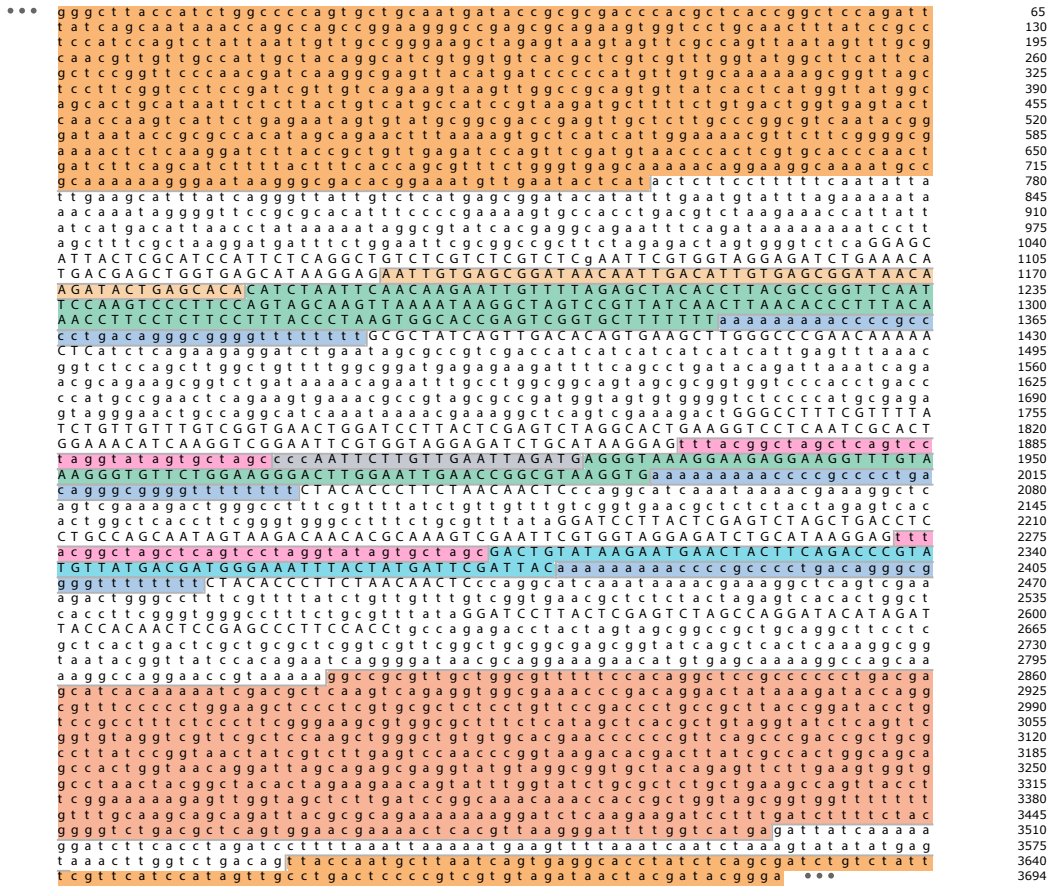Figure A.3: Annotated plasmid sequence for P1.

Figure A.4: Plasmid map for P4.

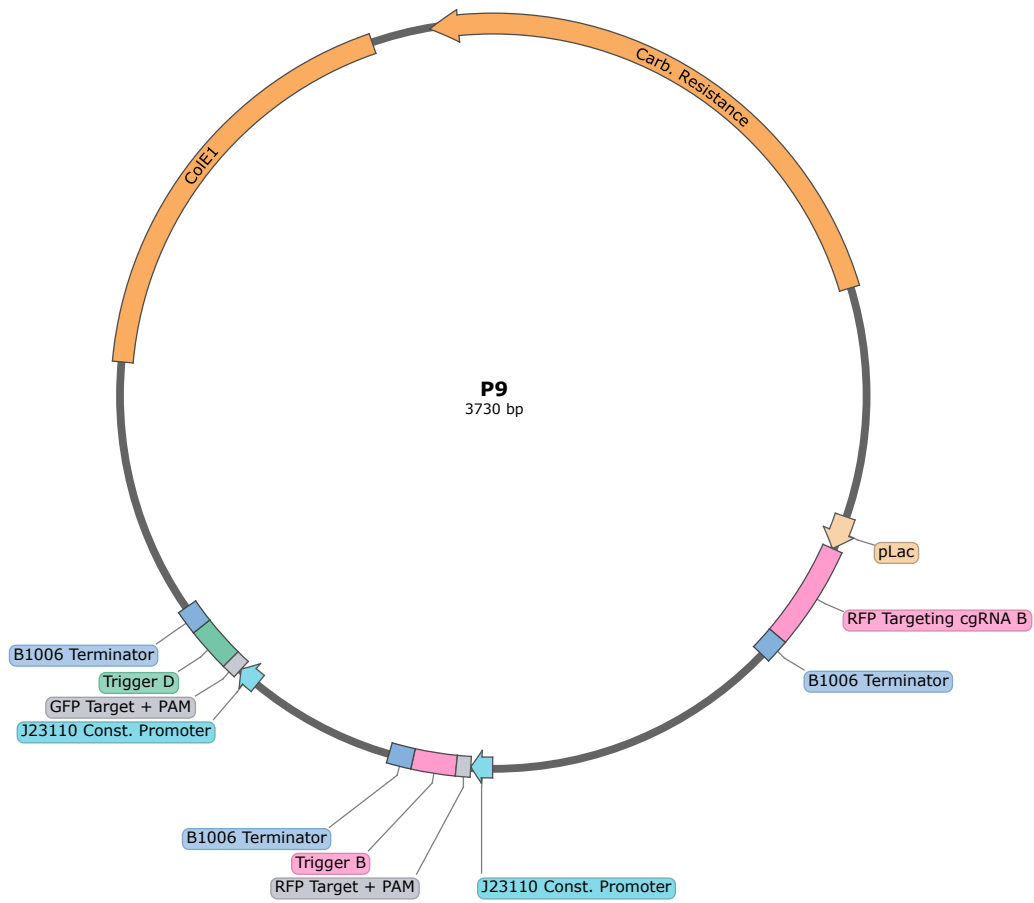Figure A.5: Annotated plasmid sequence for P4.

Figure A.6: Plasmid map for P9.

Figure A.7: Annotated plasmid sequence for P9.
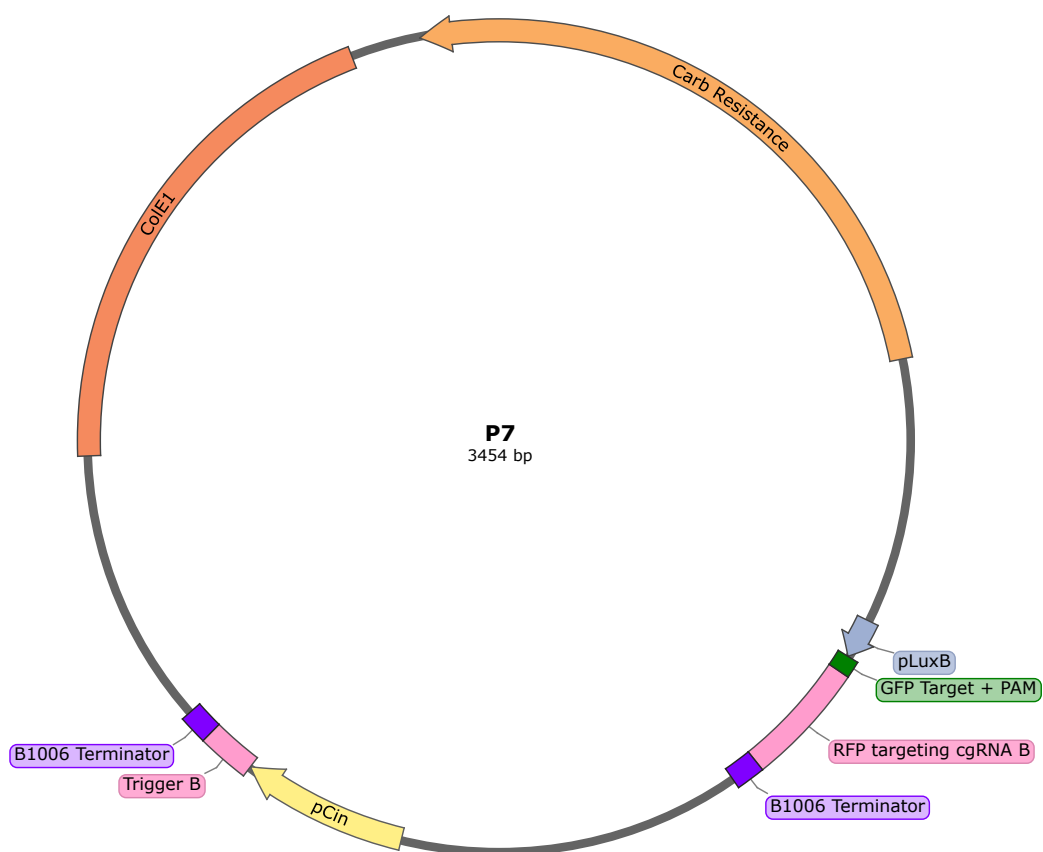
Figure A.8: Plasmid map for P7.

Figure A.9: Annotated plasmid sequence for P7.

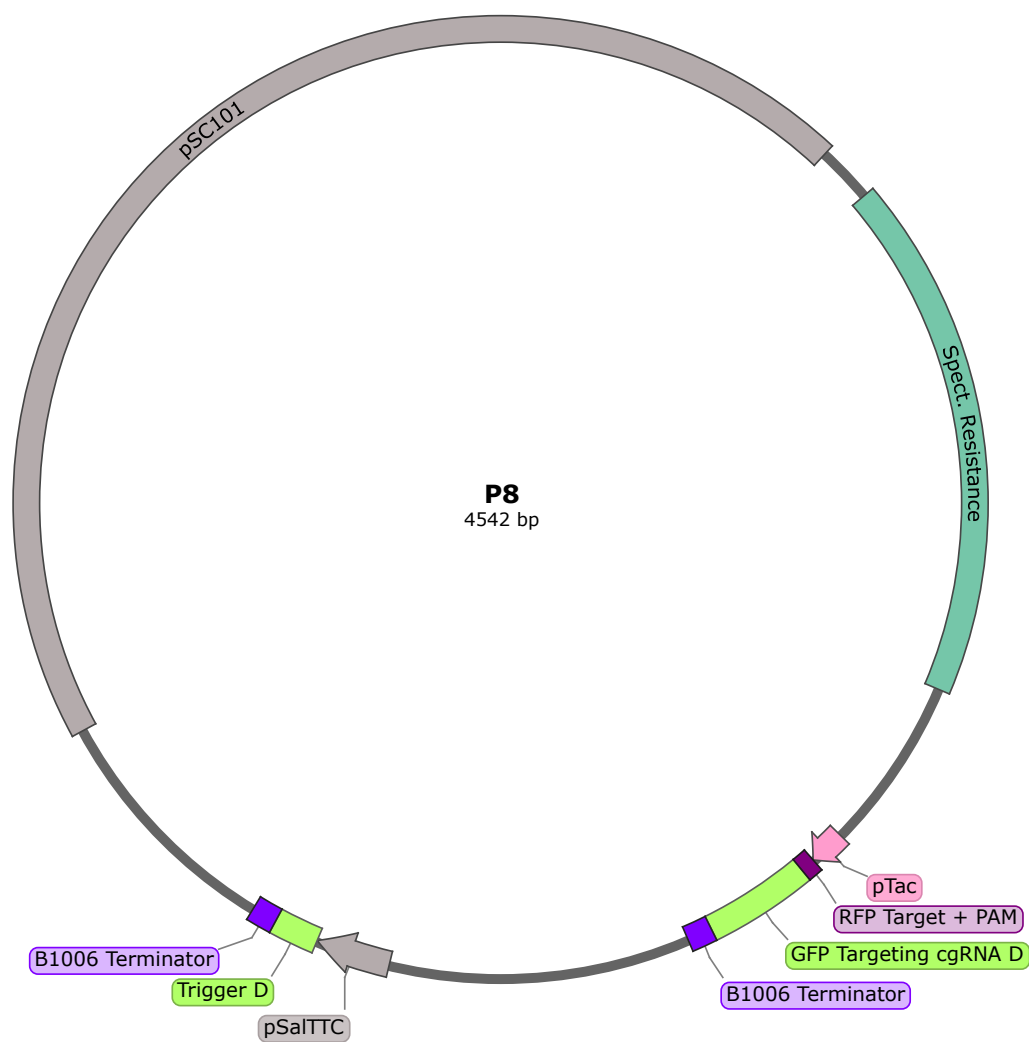Figure A.10: Plasmid map for P8.

••• GAAATTGGTGGTTGATAAGCGAGGCCGCCCGACTGATACGTTGATTTTCCAAGTTGAACTAGATA 65
GACAAATGGATCTCGTAACCGAACTTGAGAACAACCAGATAAAAATGAATGGTGACAAAATACCA 130
ACAACCATTACATCAGATTCCTACCTACGTAACGGACTAAGAAAAACACTACACGATGCTTTAAC 195
TGCAAAAATTCAGCTCACCAGTTTTGAGGCAAAATTTTTGAGTGACATGCAAAGTAAGCATGATC 260
TCAATGGTTCGTTCTCATGGCTCACGCAAAAACAACGAACCACACTAGAGAACATACTGGCTAAA 325
TACGGAAGGATCTGAGGTTCTTATGGCTCTTGTATCTATCAGTGAAGCATCAAGACTAACAAACA 390
AAAGTAGAACAACTGTTCACCGTTAGATATCAAAGGGAAAACTGTCCATATGCACAGATGAAAAC 455
GGTGTAAAAAAGATAGATACATCAGAGCTTTTACGAGTTTTTGGTGCATTTAAAGCTGTTCACCA 520
TGAACAGATCGACAATGTAACtactagagctaaatacattcaaatatctatccgctcatgagaca 585
ataaccctgataaatgcttcaataatatattgaaaaaggaagaatATGAGTGAAAAAGTGCCCGCCG 650
AGATTTCGGTGCAACTATCACAAGCACTCAACGTCATCGGGCGCCACTTGGAGTCGACGTTGCTG 715
GCCGTGCATTTGTACGGCTCCGCACTGGATGGCGGATTGAAACCGTACAGTGATATTGATTTGCT 780
GGTGACTGTAGCTGCACCGCTCAATGATGCCGTGCGGCAAGCCCTGCTCGTCGATCTCTTGGAGG 845
TTTCAGCTTCCCCTGGCCAAAACAAGGCACTCCGCGCCTTGGAAGTGACCCATCGTCGTGCACAGT 910
GACATCGTACCTTGGCGTTATCCGGCCAGGCGGGAACTGCAGTTCGGAGAGTGGCAGCGCAAAGA 975
CATCCTTGCGGGCATCTTCGAGCCCGCCACAACCGATTCTGACTTGGCGATTCTGCTAACAAAGG 1040
CAAAGCAACATAGCGTCGTCTTGGCAGGTTCAGCAGCGAAGGATCTCTTCAGCTCAGTCCCAGAA 1105
AGCGATCTATTCAAGGCACTGGCCGATACTCTGAAGCTATGGAACTCGCCGCCAGATTGGGCGGG 1170
CGATGAGCGGAATGTAGTGCTTACTTTGTCTCGTATCTGGTACACCGCAGCAACCGGCAAGATCG 1235
CGCCAAAGGATGTTGCTGCCACTTGGGCAATGGCACGCTTGCCAGCTCAACATCAGCCCATCCTG 1300
TTGAATGCCAAGCGGGCTTATCTTGGGCAAGAAGAAGATTATTTGCCCGCTCAGTGCCGGATCAGGT 1365
GGCGGCGCTCATTAAATTCGTGAAGTATGAAGCAGTTAAACTGCTTGGTGCCAGCCAATGAtaat 1430
actagctccggcaaaaaaacgggcaaggtgtcaccaccctgcccttttttctttaaaaccgaaaag 1495
attacttcgcgaaatatggcgtatcacgaggcagaatttcagataaaaaaaaatccttagctttcgc 1560
taaggatgatttctggaattcgcggccgcttctagagactagtggCATTACTCGCATCCATTCTC 1625
AGGCTGTCTCGTCTCGTCTCgAATTCGTGGTAGGAGATCTGAAACATGACGAGCTGGTGAGCATA 1690
AGGAGTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGCTCACAATTcccA 1755
GACCGCTAAACTGAAAGTTCATCTAATTCAACAAGAATTGTTTTAGAGCTACACCTTACGCCGGT 1820
TCAATTCCAAGTCCCTTCCAGTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTAACACCCT 1885
TTACAAACCTTCCTCTTCCTTTACCCTAAGTGGCACCGAGTCGGTGCTTTTTTTaaaaaaaaacc 1950
ccgcccctgacagggcgggggtttttttttGCGCTATCAGTTGACACAGTGAAGCTTGGGCCCGAAC 2015
AAAAACTCatctcagaagaggatctgaatagcgccgtcgaccatcatcatcatcatcattgagtt 2080
taaacggtctccagcttggctgtgttttggcggatgagagaagattttcagcctgatacagattaaa 2145
tcagaacgcagaagcggtctgataaaaacagaatttgcctggcggcagtagcgcggtggtcccacc 2210
tgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatg 2275
cgagagtagggaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactGGGCCTTTCG 2340
TTTTATCTGTTGTTTGTCGGTGAACTGGATCCGCACTGAAGGTCCTCAATCGCACTGGAAACATC 2405
AAGGTCGgAATTCGTGGTAGGAGATCTGCATAAGGAGGGGGCCTCGCTTGGGTTATTGCTGGTGC 2470
CCGGCCGGGCGCAATATTCATGTTGATGATTTATTTATATATCGAGTGGTGTATTTATTTATATTG 2535
TTTGCTCCGTTACCGTTATTAACAGGGTAAAGGAAGAGGAAGGTTTGTAAAGGGTGTTCTGGAAG 2600
GGACTTGGAATTGAACCGGCGTAAGGTGaaaaaaaaaaccccgcccctgacagggcgggggttttttt 2665
ttCTACACCCTTCTAACAACTCccaggcatcaaataaaacgaaaggctcagtcgaaagactgggc 2730
ctttcgttttatctgttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgg 2795
gtgggcctttctgcgtttatataGGATCCCCAGGATACATAGATTACCACAACTCCGAGCCCTTCCA 2860
CCtactagtagcggccgctgcagtccggcaaaaaaggcaaggtgtcaccaccctgcccttttttc 2925
tttaaaaccgaaaagattacttcgcgttatgcaggcttcctcgctcactgactcgctgcgctcgg 2990
tcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaattctgctagagCTGTCAGAC 3055
CAAGTTTACGAGCTCTTGGACTCCTGTTGATAGATCCAGTAATGACCTCCATCTGG 3120
ATTTGTTCAGAACGCTCGGTTGCCGCCGGGCGTTTTTTATTGGTGAGAATCCAAGCACTAGGGAC 3185
AGTAAGACGGGTAAGCCTGTTGATGATACCGCTGCCTTACTGGGTGCATTAGCCAGTCTGAATGA 3250
CCTGTCACGGGATAATCCGAAGTGGTCAGACTGGAAAATCAGAGGGCAGGAACTGCTGAACAGCA 3315
AAAAGTCAGATAGCACCACATAGCAGACCCGCCATAAAACGCCCTGAGAAGCCCGTGACGGGCTT 3380
TTCTTGTATTATGGGTAGTTTCCTTGCATGAATCCATAAAAGGCGCCTGTAGTGCCATTTACCCC 3445
CATTCACTGCCAGAGCCGTGAGCGCAGCGAACTGAATGTCACGAAAAAGACAGCGACTCAGGTGC 3510
CTGATGGTCGGAGACAAAAGGAATATTCAGCGATTTGCCCGAGCTTGCGAGGGTGCTACTTAAGC 3575
CTTTAGGGTTTTAAGGTCTGTTTTGTAGAGGAGCAAACAGCGTTTGCGACATCCTTTTGTAATAC 3640
TGCGGAACTGACTAAAGTAGTGAGTTATACACAGGGCTGGGATCTATTCTTTTTATCTTTTTTTA 3705
TTCTTTCTTTATTCTATAAATTATAACCACTTGAATATAAAAAAAAACACACAAAGGTCTAG 3770
CGGAATTTACAGAGGGTCTAGCAGAATTTACAAGTTTTCCAGCAAAGGTCTAGCAGAATTTACAG 3835
ATACCCACAACTCAAAGGAAAAGGACATGTAATTATCATTGACTAGCCCATCTCAATTGGTATAG 3900
TGATTAAAATCACCTAGACCAATTGAGATGTATGTCTGAATTAGTTGTTTTCAAAGCAAATGAAC 3965
TAGCGATTAGTCGCTATGACTTAACGGAGCATGAAACCAAGCTAATTTTATGCTGTGTGGCACTA 4030
CTCAACCCCACGATTGAAACCCTACAAGGAAAGAACGGACGGTATCGTTCACTTATAACCAATA 4095
CGCTCAGATGATGAACATCAGTAGGGAAAATGCTTATGGTGTATTAGCTAAAGCAACCAGAGAGC 4160
TGATGACGAGAACTGTGGAAATCAGGAATCCTTTGGTTAAAGGCTTTGAGATTTTCCAGTGGACA 4225
AACTATGCCAAGTTCTCAAGCGAAAAATTAGAATTAGTTTTTAGTGAAGAGATATTGCCTTATCT 4290
TTTCCAGTTAAAAAAATTCATAAAAATATAATCTGGAACATGTTAAGTCTTTTGAAAACAAATACT 4355
CTATGAGGATTTATGAGTGGTTATTAAAAGAACTAACACAAAAGAAACTCACAAGGCAAATATA 4420
GAGATTAGCCTTGATGAATTTAAGTTCATGTTAATGCTTGAAAATAACTACCATGAGTTTAAAAG 4485
GCTTAACCAATGGGTTTTGAAACCAATAAGTAAAGATTTAAACACTTACAGCAATAT ••• 4542

Figure A.11: Annotated plasmid sequence for P8.

Figure A.12: Linear insert map for genomic incorporation of RFP.

Figure A.13: Annotated linear insert sequence for genomic incorporation of RFP.

A.4 Replicates for thresholding and toggle switches

Replicates are shown for threshold and toggle switch circuit behaviors. The main text figures based on these replicates were prepared as follows:

- **Figures 2.5, 2.6, 2.7:** for single circuit in the cell threshold switches, all fluorescence values are normalized by the IPTG=0μM fluorescence value of the closed-loop circuit, before calculating the mean and standard deviation across replicates.

- **Figure 2.9:** for two threshold switches in the cell, all fluorescence values are normalized by the IPTG=0μM fluorescence value of the strain with both closed-loop circuits, before calculating the mean and standard deviation across replicates.

- **Figure 2.12:** for the toggle switch circuit, arbitrary fluorescence values were not normalized, and the mean and standard deviation were calculated from the raw AFU.

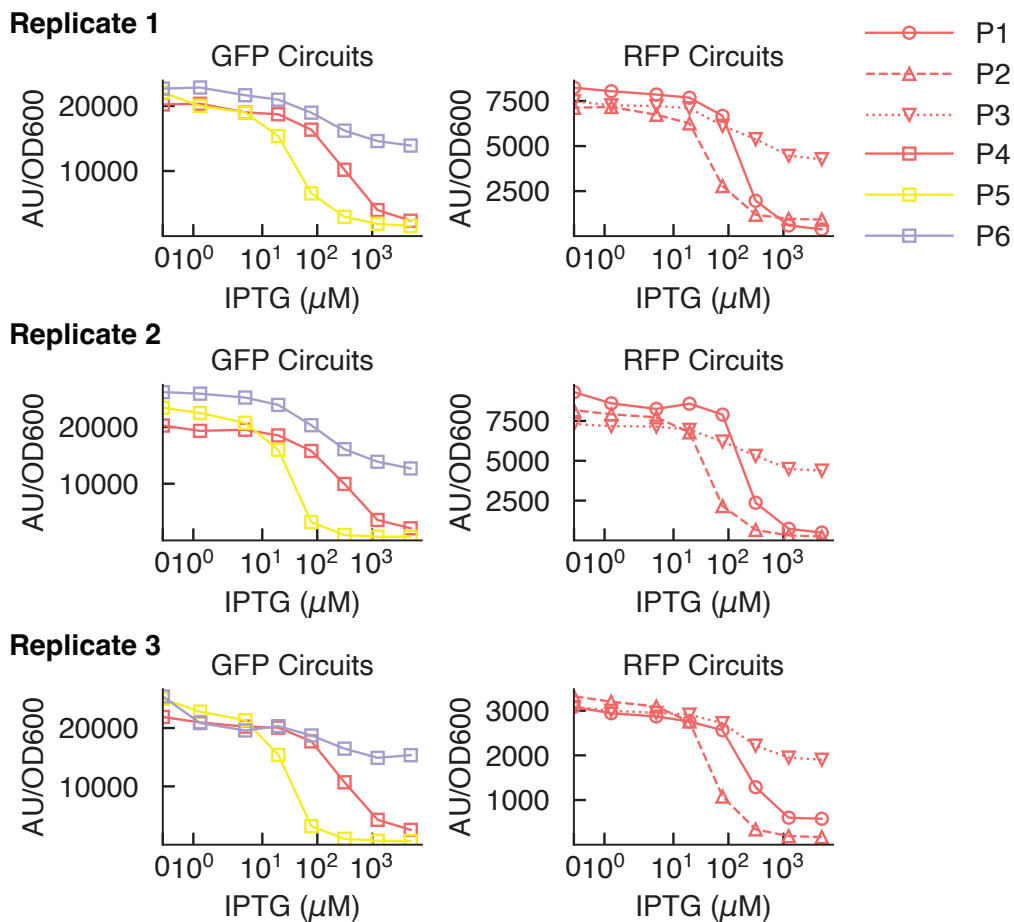**Single circuit in the cell (cf. Figure 2.5, 2.7)**

Figure A.14: Replicates for two thresholding circuits in a single *E. coli* cell (cf. Figure 2.5, 2.7) Threshold switch: IPTG induces cgRNA that transcriptionally silences trigger, which in turn hybridizes to and disables cgRNA. Open-loop controls by disrupting transcriptional inhibition (lavender scissors), which provides upper bound control; and inactivation by hybridization (yellow scissors), which provides lower bound control.

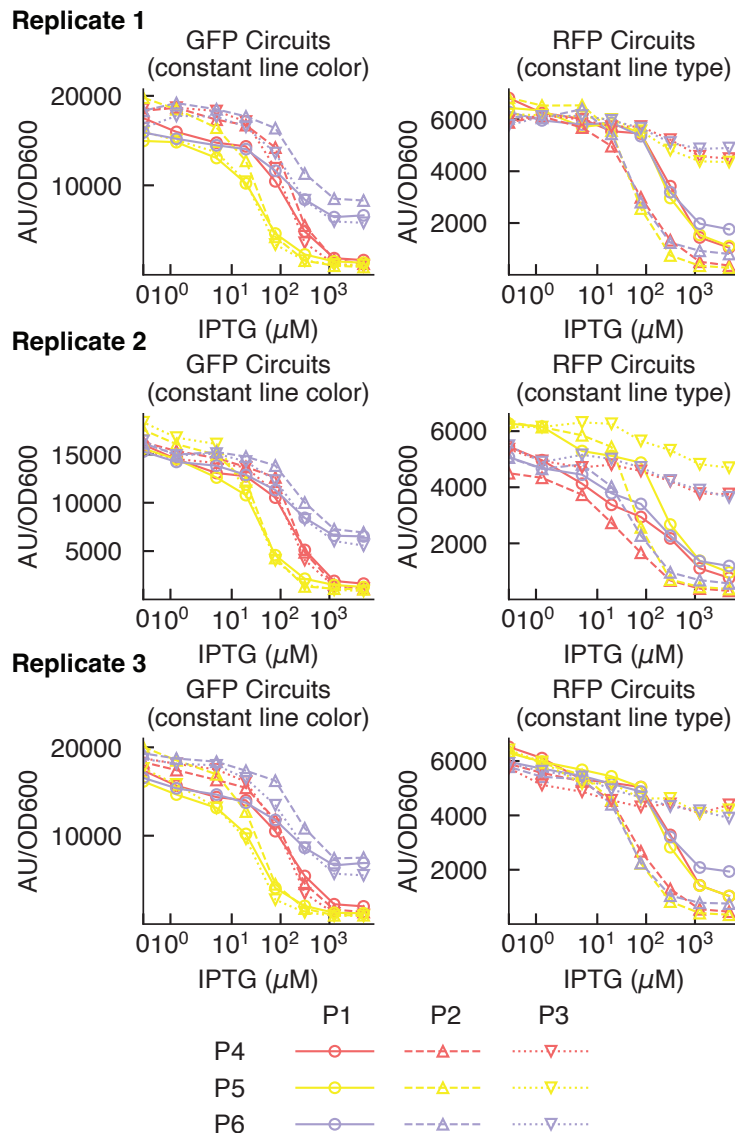**Two threshold switch circuits in the cell (cf. Figure 2.9)**

Figure A.15: Replicates for two threshold circuits in a single *E. coli* cell (cf. Figure 2.9) Threshold switch: IPTG induces cgRNA that transcriptionally silences trigger, which in turn hybridizes to and disables cgRNA. Open-loop controls by disrupting transcriptional inhibition (lavender scissors), which provides upper bound control; and inactivation by hybridization (yellow scissors), which provides lower bound control.
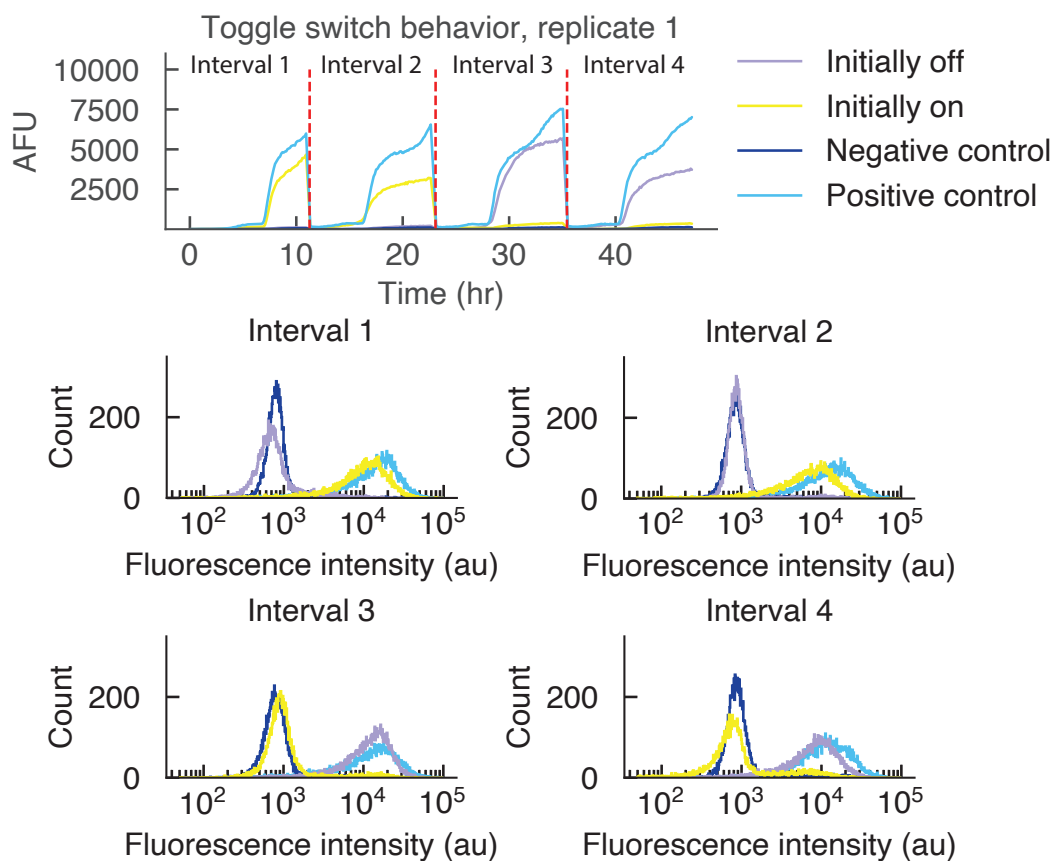
Toggle switch circuits (cf. Figure 2.12)

Figure A.16: Flow cytometry and time series replicate 1 for toggle switch circuit in *E. coli* (cf. Figure 2.12). cgRNAs silences each other to create bi-stable switch, triggers induced by small molecules disrupt cgRNA activity to actuate switch.

Figure A.17: Flow cytometry and time series replicate 2 for toggle switch circuit in *E. coli* (cf. Figure 2.12). cgRNAs silences each other to create bi-stable switch, triggers induced by small molecules disrupt cgRNA activity to actuate switch.

Figure A.18: Flow cytometry and time series replicate 3 for toggle switch circuit in *E. coli* (cf. Figure 2.12). cgRNAs silences each other to create bi-stable switch, triggers induced by small molecules disrupt cgRNA activity to actuate switch.
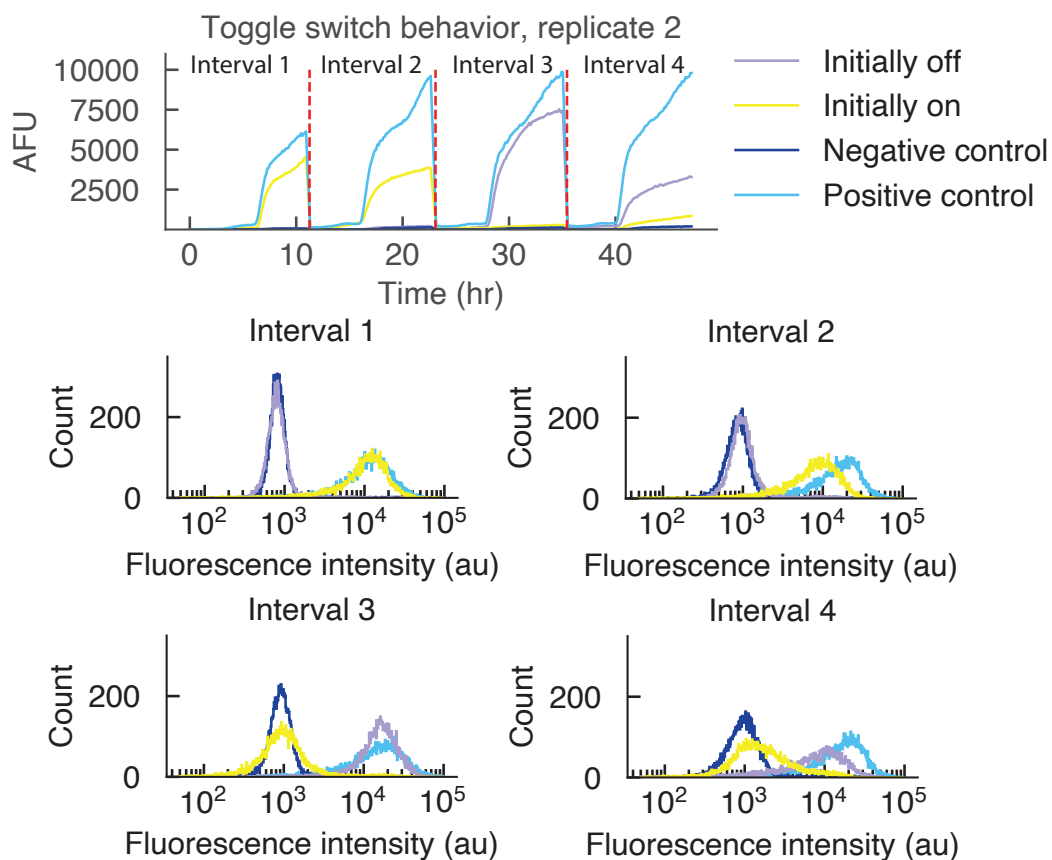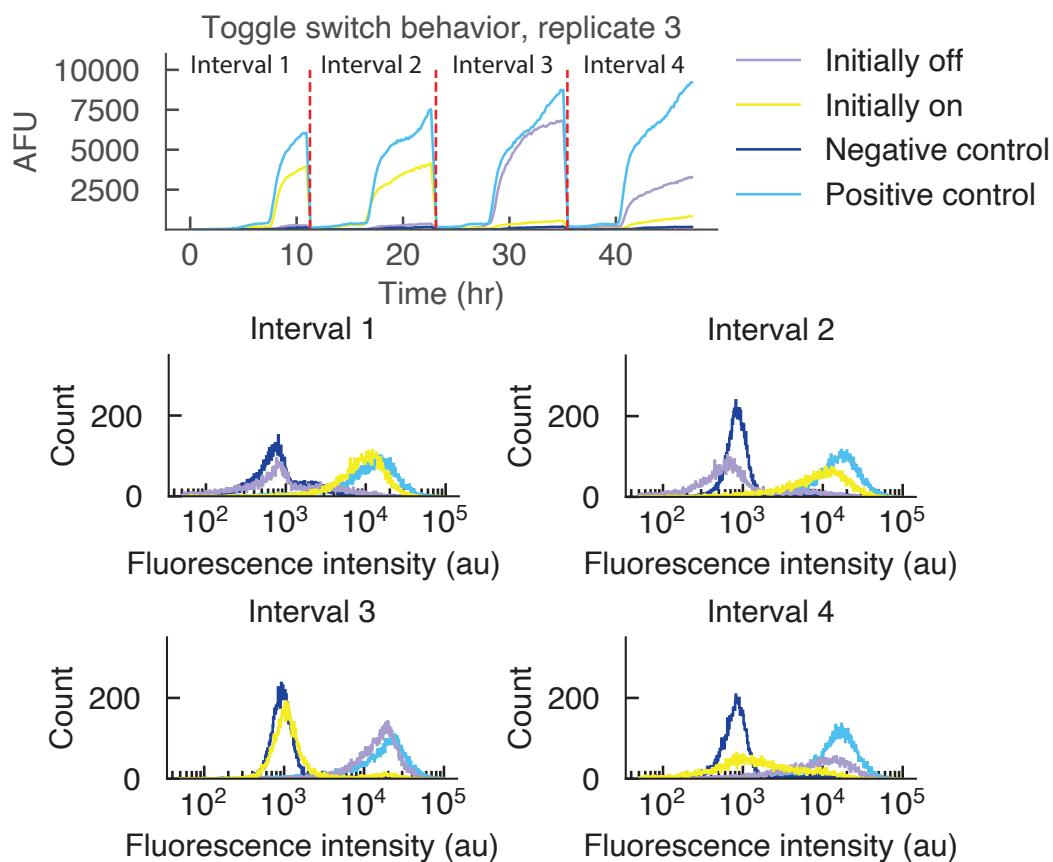
Optimization of promoter strength for thresholding switch

Here we show a series of optimization for the promoter strengths of the two constitutive promoters in the system. As the promoter strengths decrease, the two open-loop circuits collapse into one another. At the lower limit of trigger expression, trigger activity that differentiates the closed- and 2 open-loop circuits is reduced. As promoter strength increases, the gap between the two open-loop circuits grows larger. Towards the higher limit of trigger expression, the closed-loop circuit does not threshold within the range of inducers used in this experiment.

Figure A.19: Plasmid layout for testing various promoter strengths for threshold switch (cf. Figure 2.6). Description for specific plasmids available in Table A.5.

	Name	Function	Trigger 1	Trigger 2
a	**J23100 plasmids**			
	P9	Closed-loop	B	D
	P10	Open-loop by hybridization	D	D
	P11	Open-loop by transcription	D	B
b	**J23105 plasmids**			
	P12	Closed-loop	B	D
	P13	Open-loop by hybridization	D	D
	P14	Open-loop by transcription	D	B
c	**J23116 plasmids**			
	P15	Closed-loop	B	D
	P16	Open-loop by hybridization	D	D
	P17	Open-loop by transcription	D	B
d	**J23117 plasmids**			
	P18	Closed-loop	B	D
	P19	Open-loop by hybridization	D	D
	P20	Open-loop by transcription	D	B

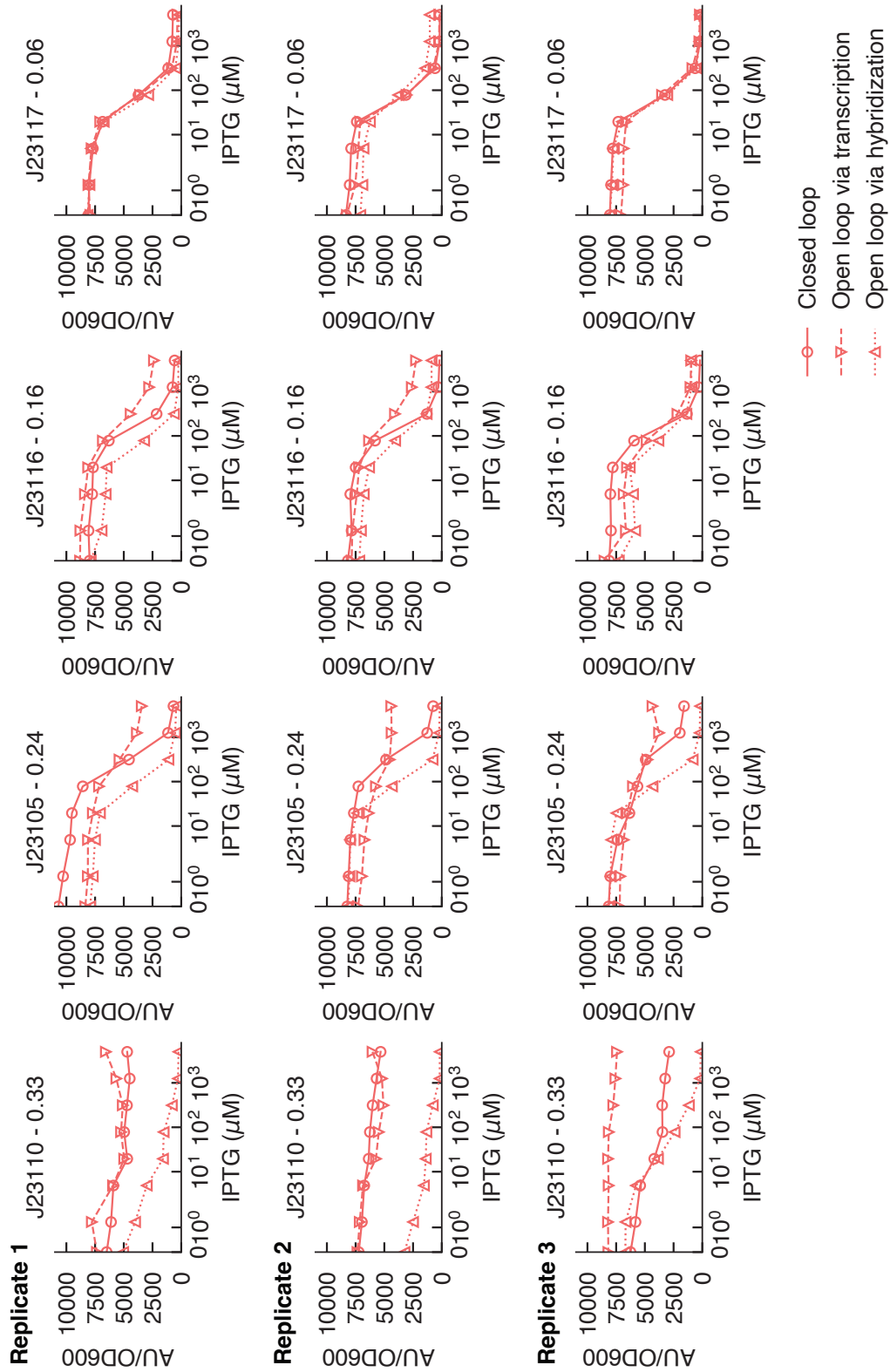Table A.5: Plasmids descriptions for threshold circuit optimization.

Figure A.20: Replicates for optimization of threshold switch circuit (cf. Figure 2.6).

A p p e n d i x B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 Materials and methods

DNA probe set purification

All probes in the SARS-CoV-2 DNA probe set (Molecular Technologies, 4014/E1052) were pooled and purified with polyacrylamide gel electrophoresis (PAGE). A 1.5 mm 20% denaturing polyacrylamide 19 cm × 20 cm gel was prepared and run in 1× tris-borate-ethylenediaminetetraacetic acid (EDTA) (TBE) (VWR, 75800-954). The gel was pre-run at 400 V for 30 min to warm the gel. The DNA probes (1000 μL, with each probe at 0.5 μM) were mixed 1:1 with denaturing loading buffer consisting of 88.8% formamide (Ambion, AM9342), 10 mM EDTA (Invitrogen, 15575-038), 2 mM Tris-HCl (pH 7.5) (Invitrogen, 15567027), 62.5 μg/mL bromophenol blue (MP Biomedicals, 193990), and 62.5 μg/mL xylene cyanol (Bio-Rad, 161-0423), and heated to 95 °C for 5 min. The entire mixture was then immediately loaded into a single well and run at 400 V for 3.5 h. The gel was unmolded onto a plastic-wrapped chromatography plate, and the DNA probe band was cut from the gel under a short wavelength UV shadow, sliced into 0.5 cm × 0.5 cm pieces, and placed in a 50 mL LoBind tube (Eppendorf, 0030122232). To the gel slices, 15 mL of 0.3 M NaCl (Promega, V4221) was added, and the tube was inverted end-over-end on a rotary wheel overnight (> 12 h). The next day, a sterile flip filter (Millipore, SCGP00525) was used to filter the liquid into a new 50 mL LoBind tube. To the filtered liquid, 35 mL 100% ethanol (Koptec, V1001) and 20 μL molecular biology grade glycogen (Thermo Scientific, R0561) were added, and the mixture was inverted several times to mix before incubating overnight at −20 °C. The 50 mL tube was then centrifuged at 4 °C for 30 min at 4.5k rcf. The supernatant was carefully removed with a pipet and discarded, and any remaining liquid was allowed to evaporate for 5 h with a Kimwipe covering the top of the tube. The pellet was resuspended in 500 μL IDTE (pH 8.0), and the total probe concentration was determined by 260 nm absorbance with a NanoDrop spectrophotometer.

Lateral flow device assembly for viral RNA detection

A 25 mm backed nitrocellulose membrane (Cytiva, FF80HP PLUS grade, 10547042) was adhered onto a polyethylene backing card (McMaster-Carr, 1441T52). A 25 mm

wicking pad (Cytiva, CF7 grade, 8117-6621) was also adhered onto the backing card with a 2 mm overlap on top of the nitrocellulose membrane. The wicking pad and nitrocellulose were then cut to size with a laser cutter (see Figure 3.5 for dimensions and Figure 3.6 for a photograph of the device). With a P2 pipette, 0.70 μL of an anti-DNA/RNA capture antibody (0.8 mg/mL) (Abcam, ab256361) was gently spotted at the test region of the nitrocellulose membrane. The nitrocellulose membrane was allowed to dry at room temperature for at least 30 min. A paper clip was added to apply pressure to the junction of the wicking pad and the nitrocellulose membrane (see Figure 3.6).

For each assay, to prepare 99 μL of Channel 1 solution without the viral sample, the following reagents were combined in nanopure water in the listed order to reach the indicated concentration: 1.5% BSA, 4$\times$ SSC, 0.05% Tween-20, 0.4 U/μL RNase inhibitor (Promega, N2615), 0.2% dextran sulfate sodium salt, and 5.4 μM of split-initiator DNA probes. For each assay, to prepare 100 μL of Channel 2 solution, 4 μL each of DIG-labeled HCR hairpins h1 and h2 (Molecular Instruments; B3-DIG) were separately snap-cooled (with h1 and h2 in separate tubes) by heating to 95 °C for 90 sec followed by cooling at room temperature for 30 min; to 92 μL Channel 2 buffer, which consisted of 0.217% dextran sulfate sodium salt in 5$\times$ SSC with 0.1% Tween-20, 4 μL each of snap-cooled DIG-labeled HCR hairpins h1 and h2 were added, and the solution was mixed by pipetting up and down and placed in a well (Channel 2 well) of a 96-well plate (Greiner Bio-One, 655901). For h1-only assays run without HCR hairpin h2, 4 μL HCR hairpin buffer was instead added (Molecular Technologies). For each assay, 4 μL CB-labeled anti-DIG reporter antibody was added to a microcentrifuge tube on ice and continuously sonicated (cycle 1.0) for 60 sec at 40% amplitude with an ultrasonic processor; 100 μL of Channel 3 solution was created by adding 4 μL freshly sonicated CB-labeled anti-DIG reporter antibody to the following reagents in nanopure water in the listed order to reach the indicated concentration: 0.35% BSA, 5$\times$ SSC, and 0.1% Tween-20, and the solution was mixed by pipetting up and down and placed in a well of a 96-well plate (Channel 3 well) right proximal to the Channel 2 well.

Performing a viral RNA detection test

Gamma-irradiated SARS-CoV-2 (BEI Resources, NR-52287) or synthetic genomic RNA from 229E (American Type Culture Collection, VR-740D) or HKU1 (American Type Culture Collection, VR-3262SD) coronaviruses was diluted as needed in nanopure water at 100$\times$ the target concentration for the test. To create 100 μL

of Channel 1 solution including the viral sample, 1μL of the diluted 100× viral stock was added to 99μL of Channel 1 solution and was mixed by pipetting up and down. Gamma-irradiated SARS-CoV-2 was quantified by BEI Resources via droplet digital PCR. This solution was incubated at 65 °C for 15 min on a heat block and transferred to a well of a 96-well plate (Channel 1 well) left-proximal to the Channel 2 well. To start the lateral flow assay, the ends of the three nitrocellulose channels (Channel 1, Channel 2, Channel 3) were simultaneously submerged in their respective wells in the 96-well plate. After 90 min, the test strip was placed in a 17-inch light tent (Angler, CT-DSLEDII) at maximum light intensity with a white background and photographed with a camera (Panasonic GH4) equipped with a 60 mm macro lens (Olympus, V312010BU000) with the following settings: 1/4000 sec shutter speed, 200 ISO, f/2.8 aperture, and neutral white balance.

Measurement of HCR signal gain using DIG-labeled HCR hairpins and CB-labeled anti-DIG reporter antibodies

To calculate the HCR amplification gain, we performed $N = 3$ replicate lateral flow assays for each of two types of experiment:

- **Amplified experiment (h1 and h2):** used both HCR hairpins (DIG-labeled h1 and DIG-labeled h2) in Channel 2 so that each HCR initiator labeling an anti-N antibody captured at the test line can trigger polymerization of a tethered DIG-decorated amplification polymer, which could then be bound by multiple CB-labeled anti-DIG reporter antibodies from Channel 3.

- **Unamplified experiment (h1 only):** used only HCR hairpin h1 (DIG-labeled h1) so that HCR polymerization cannot proceed and each HCR initiator labeling an anti-N antibody captured at the test line can bind only one DIG-labeled h1 hairpin, which could then be bound by one CB-labeled anti-DIG reporter antibody from Channel 3.

Gamma-irradiated SARS-CoV-2 was spiked into extraction buffer to a final concentration of 5,000 copies/μL and the test was run and photographed according to Section B.1. Images were converted to grayscale and the HCR amplification gain was then calculated as the ratio of amplified to unamplified intensities as described in Schulte et al. [117]. Ideally, the amplification gain would match the mean polymer length, corresponding to a situation in which a reporter antibody is binding to each hairpin within an amplification polymer.

B.2 Replicates for amplified HCR lateral flow assay

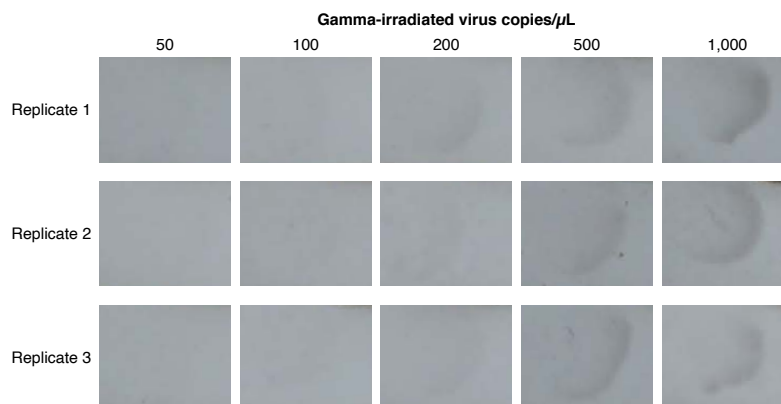Replicates for viral RNA detection of SARS-CoV-2.

Figure B.1: Sensitivity of amplified HCR lateral flow assay (cf. Figure 3.7). Following the methods of Section B.1, gamma-irradiated SARS-CoV-2 was spiked into extraction buffer with DNA probes and heated at 65 °C for 15 min; this solution was added to the Channel 1 well in a 96-well plate and the ends of the three nitrocellulose channels (Channel 1, Channel 2, Channel 3) were simultaneously submerged in their respective wells in the 96-well plate to start the test. The test line was photographed after 90 min. $N = 3$ replicate assays at each target concentration. The test line is visible in all three replicates down to a limit of detection of 200 copies/μL.
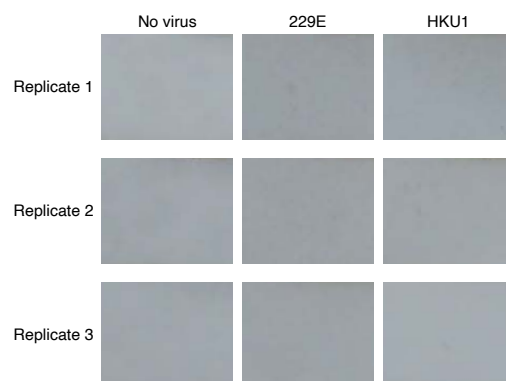
Figure B.2: Background and cross-reactivity of amplified HCR lateral flow assay (cf. Figure 3.8). Following the methods of Section B.1, no virus (to measure background) or off-target synthetic RNA genomes from coronaviruses 229E or HKU1 (to test cross-reactivity) were spiked into extraction buffer with DNA probes and heated at 65 °C for 15 min; this solution was added to the Channel 1 well in a 96-well plate and the ends of the three nitrocellulose channels (Channel 1, Channel 2, Channel 3) were simultaneously submerged in their respective wells in the 96-well plate to start the test. For the cross-reactivity tests, the off-target viral RNA genomes were spiked in at high concentration (7,200 copies/μL for 229E and 10,000 copies/μL for HKU1). The test line was photographed after 90 min. $N = 3$ replicate assays at each target concentration. For each target type, no staining was visible at the test line for all three replicates, indicating that there is no visible background when SARS-CoV-2 is absent, and no visible cross-reactivity with off-target 229E or HKU1 coronavirus RNA genomes.
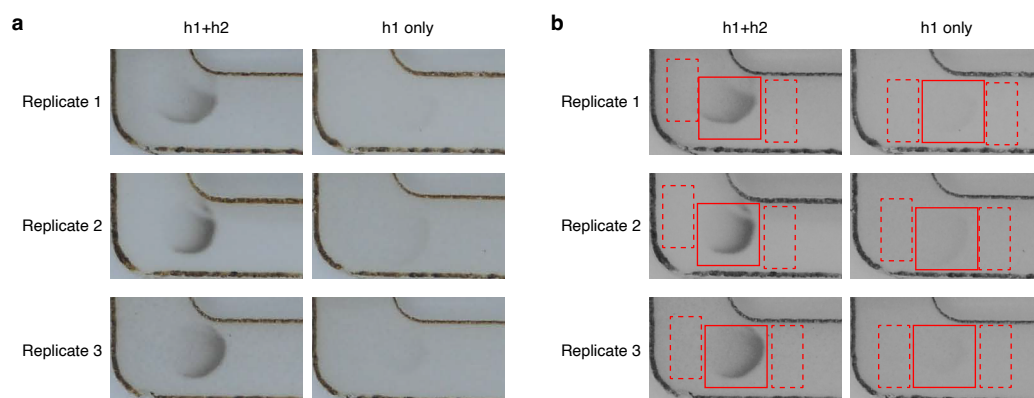
Figure B.3: Measurement of HCR amplification gain in the context of a lateral flow assay for viral RNA detection (cf. Figure 3.9). Following the methods of Section B.1, two types of experiment are compared: amplified assays (using both h1 and h2, so that HCR polymerization can proceed) vs unamplified assays (using h1 only, permitting only a single h1 binding event with no polymerization due to the absence of h2). Gamma-irradiated virus was spiked into extraction buffer at 5000 copies/μL for all tests. $N = 3$ replicate assays for each condition. (a) Raw images. (b) Images after conversion to grayscale. Quantitative image analysis following the methods of Schulte et al. [117] using the depicted signal boxes (solid boundary) and background boxes (dashed boundary).