Utilizing RNA aptamers for small molecule biosensing in engineered biosynthetic pathways

Chuhern Hwang

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

Doctor of Philosophy

University of Washington
2019

Reading Committee:
James Carothers, Chair
Herbert Sauro
Wendy Thomas

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Utilizing RNA aptamers for small molecule biosensing in engineered biosynthetic pathways

Chuhern Hwang

Chair of Supervisory Committee:
Assistant Professor James Carothers
Department of Chemical Engineering

Chemical production from renewable feedstocks using engineered microbes will play an important role in advancing sustainable technological solutions to climate change and other environmental challenges. Biosensors are already playing a critical role in production screening and dynamic genetic control to reduce the time and costs and increase titers in engineered biosynthetic pathways. RNA aptamers are an increasingly utilized tool for metabolite biosensing but matching the sensitivity of RNA aptamer-based devices to the titers from engineered pathways is challenging and unpredictable. This thesis addresses that problem, modeling the energetic contributions of metabolite characteristics towards RNA aptamer binding affinities to predict pathways and molecules more amenable to RNA aptamer biosensing. We present tools and approaches for selecting RNA aptamers and using those aptamers to build functional, responsive RNA biosensors. This thesis details a label-free *in vitro* selection for an RNA aptamer binding the metabolite *p*-aminocinnamic acid. This work presents, for the first time, an approach to predict binding affinities of RNA aptamers towards small molecule targets from the chemical properties of the small molecules. We demonstrate the utility of kinetic aptamer ribosensors in

reporting titers of the small molecule *p*-aminophenylalanine in supernatants from engineered cells. This thesis also details the first demonstration of engineering metastable RNA folding states to encode time using multistate co-transcriptional RNA design. Collectively, this work provides tools for assessing the suitability of RNA aptamers for sensing applications in metabolic engineering while simplifying the process of *in vitro* selections for small molecule aptamers and providing a battery of tools for tuning the sensitivities of kinetic co-transcriptional RNA biosensors.

Acknowledgments

The six-year journey to this Ph.D. was far from a solitary one and would not have been possible without the guidance and support from my mentors, friends, and family.

First and foremost, I would not be here without my adviser, Dr. James Carothers, who gave me ample latitude throughout my time in graduate school to explore even ideas that he was (usually correctly) highly skeptical about. I would like to thank the members of my thesis committee, Dr. Wendy Thomas, Dr. Herbert Sauro, Dr. Cole DeForest, and Dr. Luis Ceze, as well for their feedback and guidance on my dissertation work.

I would like to thank Dr. Cassandra Burke for her constant and invaluable mentorship during my time in the Carothers Lab. Without her constant sunny guidance in the lab I'm certain I would not be here today. I gained critical insights from my conversations with David Sparkman-Yager, whether on RNA folding or basketball. I appreciated and learned much from Willy Voje's thoughts on grad school and RNA. It was a pleasure and honor to learn from and work with Jason Stevens, Rodrigo Correa, Jason Fontana, Ian Faulkner, Ben Tickman, Pramod Chavali, Chenggang Xi, Neel Shah, Paige Bennett, Austin Wright-Pettibone, Trenton Grossfeld, and other members of the Carothers Lab.

I'm deeply grateful to Sifang Chen, Randolph Lopez, and Sundipta Rao for their support, both scientific and emotional, throughout my time in graduate school. Their friendship has been a cornerstone of my life and I will always cherish the times we spent together. I'm also grateful for my continued friendship with Alex Nguyen; although we no longer pull all-nighters together working on undergrad lab reports, Alex has been no less supportive during my time in grad

school. I am profoundly thankful to Anna Li for her patience and emotional support, providing motivation and uplifting me to complete my thesis work.

There are far too many people whose friendships and scientific support helped guide me through my six years here and I appreciate every act of goodwill and scientific collaboration that I've received from friends and colleagues regardless of whether I've mentioned them here. I do want to thank Alex Rosenberg, Elizabeth Heath, Sumit and Ladan Mukherjee, Leandra Brettner, Gary Liu, Alberto Carignano, Bob Lamm, Nandana Rao, Charlie Roco, Max Darnell, Paul Sample, and Leli Amimeur for the enriching interactions they provided throughout grad school.

Lastly, I wouldn't have made it through grad school without the sustained support of my parents and my sister. I would also like to thank my uncles and aunts who made the trip to Seattle over the past six years and expressed their support, and both of my grandmothers, whose extraordinarily inspirational lives are a primary motivation in my life.

Table of Contents

Chapter 1. Introduction and background

1.1 Motivation

The growth of industrial civilization in the past two and a half centuries has powered an unprecedented improvement in standard of living for humans all over the planet but has done so in a resource-intensive way that has had severe impacts on the ecology of this planet. Our dependence on material goods and fuels that are chemically derived from petroleum is both ultimately unsustainable due to the limits in fossil fuel supplies and a major contributor to anthropogenic climate change. The gasoline that we use to fuel our cars; diesel for trucks; kerosene for aircraft; complex pharmaceuticals for human health; high-value polymers for industrial applications; consumer plastics found in our everyday lives; these are only a small sampling of how petroleum-based products dominate our lives.

The twenty-first century has opened with a broad consensus and acknowledgment that fossil fuel use is unsustainable, and that continued use will contribute more to increasing levels of atmospheric carbon dioxide and ultimately exacerbate the effects of anthropogenic climate change. This dissertation does not aim to enumerate the effects of climate change in detail, but in brief, climate change is expected to affect quality of human life in many ways, increasing the chance for severe weather events across the world, bringing sea level rise that will affect many of the most densely populated regions in the world, increasing the incidence of many diseases, and affecting agricultural yields[1]. Such broad-ranging negative effects may largely offset many of the quality of life improvements that have resulted from these technologies and products. The Intergovernmental Panel on Climate Change (IPCC) has projected ranges of future outcomes

based on ranges of emission targets; international agreements such as the Paris Agreement on Climate Change aim to unite nations in implementing emissions reductions that will mitigate the worst effects of climate change. Greenhouse gas emissions result from a broad range of human activities, but about 29% of emissions in the United States result from transportation and another 22% from industrial activities[2]. By reducing use of petrochemicals in these sectors, we can reduce the contribution of these sectors to greenhouse gas emissions and make the emissions reductions necessary for mitigation of extreme climate effects a more achievable goal.

The field of metabolic engineering is exploring bio-based alternatives to existing methods of chemical production, using programmed cellular factories to enzymatically convert renewable or waste feedstocks into useful chemical products. Companies and governmental and academic institutions have developed engineered metabolic pathways in microbes for pharmaceuticals such as artemisinin; food additives such as vanillin; and biofuels such as farnesene[3-5]. These avenues of work look to displace petroleum as the source material for the fuel and material goods that we use, instead relying on agricultural feedstocks or waste that can have a smaller carbon footprint.

By finding ways to lower costs, increase titers, or improve strain optimization in metabolic engineering, we improve the competitiveness of cellular refineries against petroleum refineries and we speed the development of scaled-up, cost-competitive chemical production approaches that are more sustainable and have a reduced carbon footprint. There are few issues more pressing than climate change; the promise of metabolic engineering as an avenue to help bring down anthropogenic greenhouse gas emissions motivates the work presented in this thesis.

1.2 The role of sensors and genetic controllers in metabolic engineering

The core purpose of metabolic engineering is to build cellular factories that can implement specific metabolic pathways for producing useful and valuable chemical compounds. This has already been successfully applied to examples such as insulin[6], artemisinin[4], taxol[7,8], and biofuels[3] and is the subject of active work in a variety of pathways in both academic institutions and industry. Cellular metabolism is a powerful mechanism through which we can increase efficiency and yields of natural products, build *de novo* metabolic pathways, and access molecules that are outside the current reach of synthetic chemistry[9]. Designing and optimizing biosynthetic pathways is difficult: context-dependent enzymes and genetic parts, nonlinear responses to metabolic strain, and toxic metabolic intermediates are just some of the many problems that plague pathway engineering work[10,11].

Screening engineered cells for optimized titers and engineered metabolite-responsive dynamic genetic control are two methods to tackle these problems. High-throughput screening methods are critical for metabolic engineering work, whether for measuring outputs from directed evolution approaches or for rational pathway design[12]. In dynamic genetic control, the expression of pathway enzymes can be selectively governed by the concentration of specific metabolites, linking metabolic flux to small molecule presence.

Bioreceptors that can detect metabolites are critical for both high-throughput screens and dynamic genetic control and thus to advances in the field of metabolic engineering. Analytical chemistry approaches such as HPLC, NMR, or mass spectrometry allow for metabolite quantification but are generally poorly suited for high-throughput screens, as such approaches are both slow and expensive. Dynamic genetic control requires sensors that can interface with

gene expression machinery, which for now still means a reliance on biological sensing mechanisms.

A fundamental concern with any sensor is sensitivity. For any given application, a sensor's utility depends on its transfer function covering the relevant range of the measured property. In other words, sensors are only useful if they can demonstrate differences in output for differences in the measured property that need to be distinguished for the specific application. In the context of biosensors, this means that a sensor that is sensitive to millimolar concentrations of a target molecule lacks utility in the context of an application requiring detection at the nanomolar range. This is an important point, as biosensors require bioreceptors such as proteins and nucleic acid aptamers, and the sensitivity achievable from sensors constructed using such bioreceptors ultimately depends on the dissociation constant (K_D) between the bioreceptor and the molecular target. Even when a bioreceptor exists for a given molecular target, a mismatch between the binding affinity of the bioreceptor for its target and the concentrations of the target that must be detected for a given application may render the biosensor solution useless in that context. In this thesis, we will explore using RNA aptamers for sensing in metabolic engineering contexts.

1.3 Engineered metabolic pathway for *p*-aminostyrene

We are developing metabolite-responsive genetic devices and readouts in the context of broader work on engineering a metabolic pathway for biosynthesis of *p*-aminostyrene (*p*-AS) in *E. coli*, an aromatic monomer with biomedical and photonics applications. The *p*-AS pathway is built off the existing chorismate synthesis pathway in *E. coli*, which is used for production of aromatic amino acids. Heterologous enzymes convert chorismate stepwise into two metabolic intermediates, *p*-aminophenylalanine (*p*-AF) and *p*-aminocinnamic acid (*p*-ACA) (Fig. 1.1)[13,14]. Intermediates can feed into pathways for production of other useful chemical products: *p*-AF is a starting point for synthesizing the pristinamycin family of antibiotics while *p*-ACA is useful for making cinnamic acids[15]. However, this engineered pathway also presents a number of problems: *p*-ACA is a cytotoxic intermediate and overexpression of pathway-related enzymes can create a heavy metabolic burden[14]. This makes the *p*-AS pathway an ideal pathway both to demonstrate the utility of RNA aptamer-based devices in the context of biosensing readouts and genetic control and to find a path to producing a useful and valuable molecule through the framework of metabolic engineering.

Mechanistic modeling of the pathway suggests that dynamic control in the form of metabolite-responsive gene regulation is important for optimizing titers[14]. An RNA aptamer has been reported for *p*-aminophenylalanine (*p*-AF) in previous work and has been incorporated into devices for regulating gene expression and measuring production titers[16-18].

There have not yet been demonstrations of *p*-ACA production from glucose, although *p*-ACA has been produced in E. coli from spiked *p*-AF from cells heterogeneously expressing PAL2[19].

In ongoing work in the Carothers Research Group, the relation between expression levels of heterologous and native genes and production of *p*-AF and *p*-ACA in engineered *E. coli* is being mapped for optimization of pathway strains. Titers of about 200 mg/L have been achieved for *p*-AF from glucose in *E. coli*[20]. *p*-AF and *p*-ACA production is currently quantified using high-performance liquid chromatography (HPLC), a sensitive but low-throughput approach for quantifying metabolite concentration. Current HPLC workflows require 20 minutes for quantification of a single sample, which makes quantifying larger numbers of strains time-consuming and impractical. High throughput screening methods for sensitive detection of intermediate metabolites are necessary for large scale mapping of pathway combinatorial expression space.

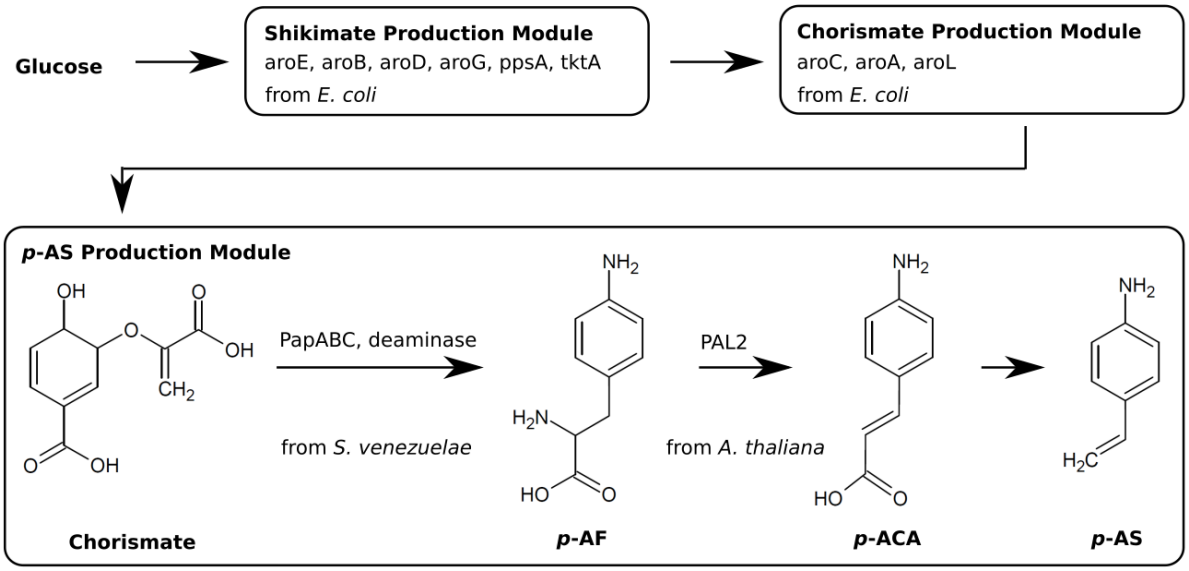


Figure 1.1. Biosynthesis of *p*-AS from glucose. Chorismate is produced endogenously in *E. coli*; the PapABC operon and PAL2 are heterogeneously expressed as part of a five gene pathway for enzymatically converting chorismate into *p*-ACA. Conversion of *p*-ACA to *p*-AS has not yet been demonstrated.

1.4 *In vitro* selection of RNA aptamers

Aptamers are nucleic acid structures capable of binding a target ligand. Although central dogma characterization seems to limit RNA to a passive information-carrying role, in biology, RNA folds into complex, functional three-dimensional structures. Aptamers are one such structure that interact with targets as diverse as small molecules, proteins, and cells through non-covalent binding[21-23].

Both synthetic and natural aptamers exist: natural aptamers are often found as components of riboswitches, a biosensing genetic control device that occurs in bacteria[24]. Synthetic aptamers, which are generated from *in vitro* selection, have been engineered into metabolite-responsive control devices and have been used for readouts of ligand presence[17,25,26]. Aptamers have been selected to bind a wide range of targets, including small molecules, proteins and cells. To date, about 60 small molecules have characterized RNA aptamers generated by *in vitro* selection[27-29].

In vitro selection, or the systematic evolution of ligands by exponential enrichment (SELEX), was pioneered in tandem by Tuerk and Gold and Ellington and Szostak as an approach for generating small molecule or protein-binding RNAs from diverse sequence spaces[21,22]. Typically, RNA aptamers are enriched from large nucleic acid pools of 10^{12} to 10^{15} random sequences through an iterative affinity chromatography approach. Chemistries for immobilizing the target molecule are necessary for the standard implementations of this approach, although newer developments such as capillary electrophoresis SELEX make immobilization unnecessary for larger targets such as proteins[30]. An initial diverse, randomized DNA pool is generated and then *in vitro* transcribed into RNA. RNA sequences capable of binding a given target are

partitioned through exposure to the immobilized target and are then typically recovered by elution with the target ligand in solution (Fig. 1.2)[31]. Recovered sequences are reverse transcribed, amplified through PCR and transcription, and are then subject again to affinity chromatography with the immobilized target. The progress and success of an aptamer selection is monitored by quantifying the relative amount of RNA that is eluted with the ligand after each round of selection; enrichment of binders is reflected as an increase in the fraction of recovered RNA after affinity chromatography. After enrichment saturates, the pool is sequenced; typically, putative binders are further optimized by reselection using doped pools[16]. Further sequencing analysis reveals base conservation at different positions on the aptamer and provides structural insights of the binder.

In vitro selection is a time-consuming process. Due to the many rounds of selection necessary and due to the lack of feedback on enrichment during the process, selections typically take multiple months. Analyses of *in vitro* selection conditions show that the median RNA aptamer selection requires 10 rounds, with some reported RNA aptamer selections taking as many as 29 rounds[32]. Due to the financial and time costs of sequencing, sequencing is generally run at the conclusion of a selection rather than after each round, which results in minimal feedback on the enrichment of sequences over the course of the selection. Although the overall success rate of running *in vitro* selections has not been characterized, anecdotes in literature sources suggest that a majority of selections are unsuccessful[27].

Multiple efforts are underway to cut the time investment necessary for selections: particle display methods reduce the number of rounds of selection necessary by employing FACS and CE SELEX attempts for more efficient partitioning and avoidance of target immobilization through capillary electrophoresis, while the MEDUSA and RAPID techniques parallelize the

process of multiple selections to increase throughput and reduce the time necessary for selections by performing multiple partitions per amplification step[30,33-35]. Other modifications aim to avoid the need for radiolabeled RNA, instead employing tracking methods such as surface plasmon resonance[36]. Many of these approaches still require expensive and specialized equipment or radiolabeling of RNA, which may be inaccessible to many labs interested in generating aptamers for metabolic engineering or biomedical applications. In other cases, solutions may be applicable to SELEX for protein targets but may be inappropriate for use with selections against small molecules; for instance, CE SELEX can use capillary electrophoresis to separate protein-bound RNAs from free RNA easily due to the notable size difference between those species, but is much more difficult to implement when separating RNA sequences that are each around 30 kDa from small molecules with molecular weights that are often just a few hundred Daltons.

RNA aptamers are a solution well-suited for many biosensing applications, as they are comparably easier to generate than protein biosensors for molecules lacking known sensors and they can be incorporated into genetic control and biosensing readout devices through design methods utilizing powerful RNA secondary structure prediction tools.

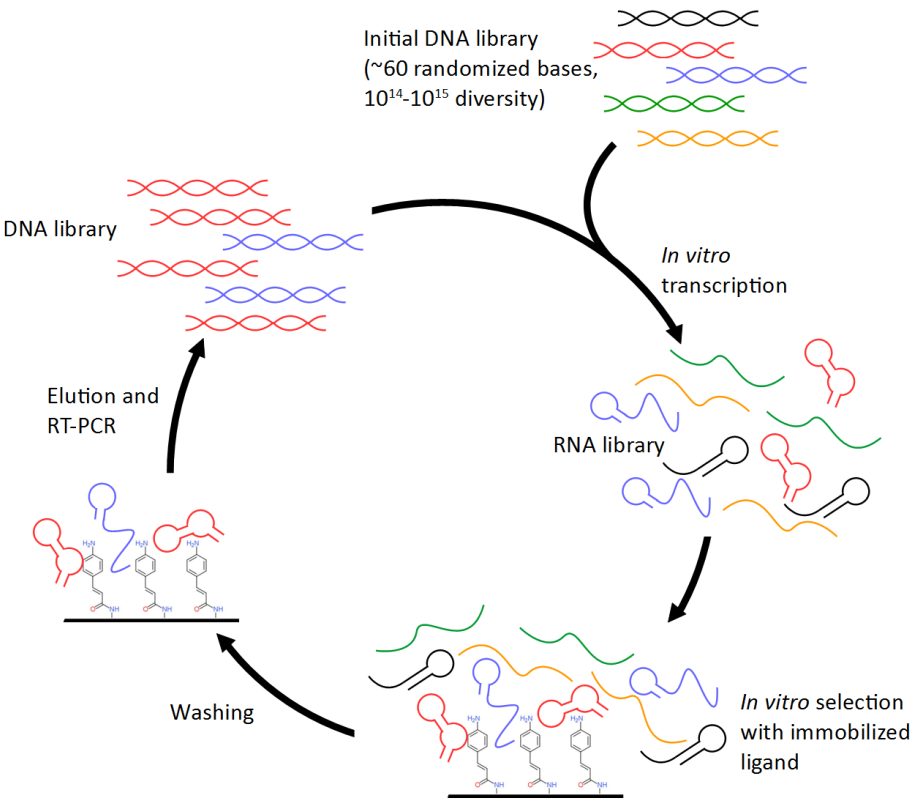


Figure 1.2. The iterative process of *in vitro* selection for RNA aptamers. Large, diverse DNA pools are transcribed into RNA and partitioned using affinity chromatography with an immobilized target. Surviving sequences are eluted with the target molecule in solution and reverse transcribed and amplified to restart the cycle, enriching the sequences that can bind to the target.

1.5 Aptamer ribosensors for high-throughput small molecule detection

RNA biosensing devices that act through kinetic control mechanisms are common in biology and can generate highly sensitive and specific responses[37-39]. The function of kinetic devices is improved by their irreversibility and dependence on the reaction rates rather than on limits occurring from reaction thermodynamics. In kinetically controlled devices, the time window that the aptamer is available for binding governs the sensitivity of the device. It has been shown that the *E. coli* thiamine pyrophosphate (TPP) riboswitch is capable of detecting progressively lower concentrations as the binding time window of the device is progressively extended[40]. Ability to tune the time window for binding can thus allow for tuning the sensitivity of aptamer-based kinetic devices. Association kinetics also play a major role in the sensitivity of kinetic devices: for any given length of binding time, slower on rates indicate lower sensitivity[40].

Recent work by Burke demonstrates a method for designing kinetically controlled RNA aptamer ribosensor devices that initiate a toehold-mediated strand displacement (TMSD) output circuit with fluorophore reporters[18]. In this approach, RNA aptamer ribosensors are engineered to fold into a co-transcriptionally available ligand-competent state in which the aptamer is capable of binding available target. Upon transcription of an actuation domain to activate a TMSD output, either the device is stabilized by ligand binding, initiating a TMSD reaction output, or the device refolds into a thermodynamically favored off state in which both the aptamer and output toehold are sequestered (Fig. 1.3). Timer sequences sandwiched between the aptamer and the actuator elongate total transcription time of the device, providing a time window for ligand binding prior to transcription of the actuator. Secondary structure prediction tools, including the

inverse-folding package MODENA and the Vienna RNAfold package, were used to assess multi-state designs[41,42].

The relationship between the EC_{50} of a kinetic device and the binding time for that device can be derived from the first order rate equation and expressed in the form:

$$EC_{50} = -\frac{ln(0.5) - k_{off} \times t_{bind}}{k_{on} \times t_{bind}} \quad \text{Eq. 1.1}$$

Where t_{bind} is the time when the aptamer is available to interact with ligand. Tuning t_{bind} thus allows us to tune the sensitivity of a kinetic device: in the context of kinetic aptamer ribosensor design, this tuning is most easily done by altering the length of the timer sequence, which alters the amount of time necessary to transcribe the sequence that follows the aptamer. Additional strategies on extending t_{bind} and thus on improving kinetic device sensitivity include slowing transcription speed and engineering metastable folding states, both of which we will address in this dissertation.

Limiting aptamer-ligand interactions to a kinetic, co-transcriptional binding window requires engineering RNAs that will fold reliably through a series of structure states for binding and actuation. Folding transitions between structure states for kinetic aptamer ribosensors are governed by free energy barriers separating those states; RNA folding tools that predict the height of those energy barriers separating distinct structure states have already been utilized to engineer functional RNAs[43]. The transition rates between structure states for aptamer ribosensors were engineered by computationally screening for and selecting sequences exhibiting appropriate free energy barriers[18].

Toehold-mediated strand displacement reactions (TMSD) are a tunable and specific mechanism that allow for construction of DNA circuits[44]. In TMSD, single-stranded nucleic acid inputs displace output strands from double-stranded DNA complexes by binding to unpaired

toeholds in the double-stranded complexes and competing for hybridization with the output strand through branch migration[45]. Multiple strategies for TMSD readout have been developed; existing aptamer ribosensors displace fluorophore-conjugated strands from quencher strands to generate a fluorescent output in response to kinetically controlled ligand-induced ribosensor activation.

Aptamer ribosensors were developed for *p*-AF and tyrosine *in vitro* selected aptamers[16,46]. A collection of *p*-AF aptamer ribosensors exhibited titratable ligand-dependent signal *in vitro*, providing a demonstration of this platform (Fig. 1.4). The application of these sensors for detecting and quantifying metabolites of interest from engineered cells could make this platform a valuable and efficient method for high throughput screening. Ribosensor assay quantification would allow for processing of up to 32 samples in an hour, representing approximately a 10-fold increase in throughput efficiency over existing HPLC approaches.

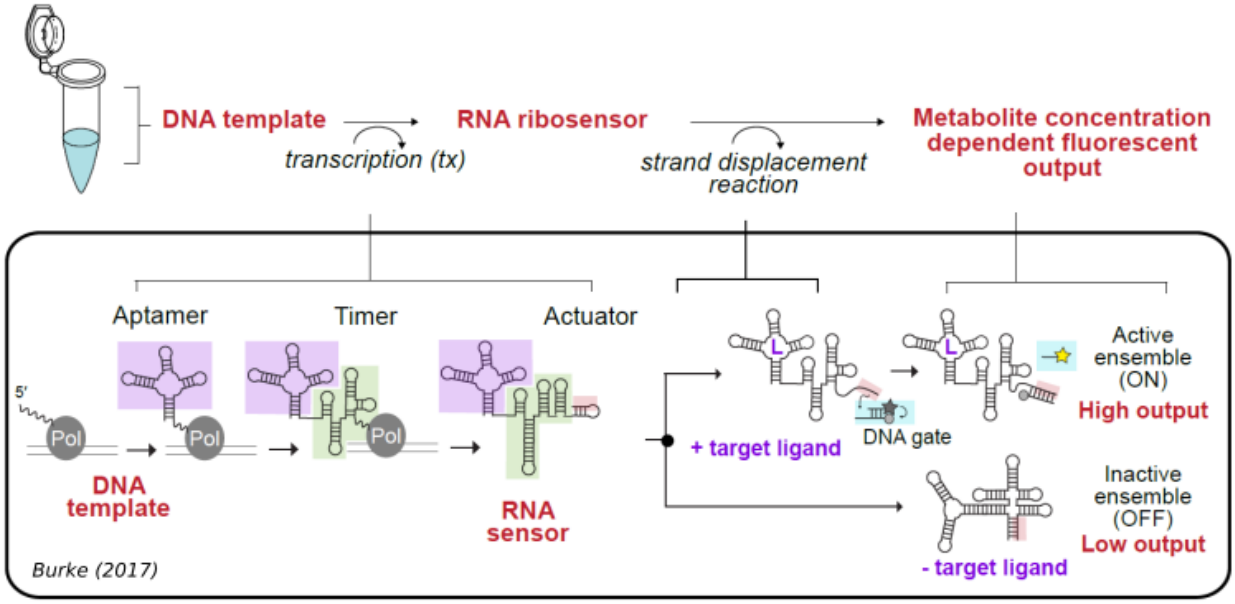


Figure 1.3. Kinetically controlled co-transcriptional aptamer ribosensor devices that deliver metabolite-dependent fluorescent outputs. A ligand-competent aptamer structure state is present during transcription, creating a time window for ligand binding during transcription. At the end

of transcription, the ribosensor is either stabilized by ligand binding and initiates a strand displacement reaction to trigger a fluorescent output, or the ribosensor undergoes a structural rearrangement to fold into an off state absent ligand binding. Adapted with permission from Burke (2017).

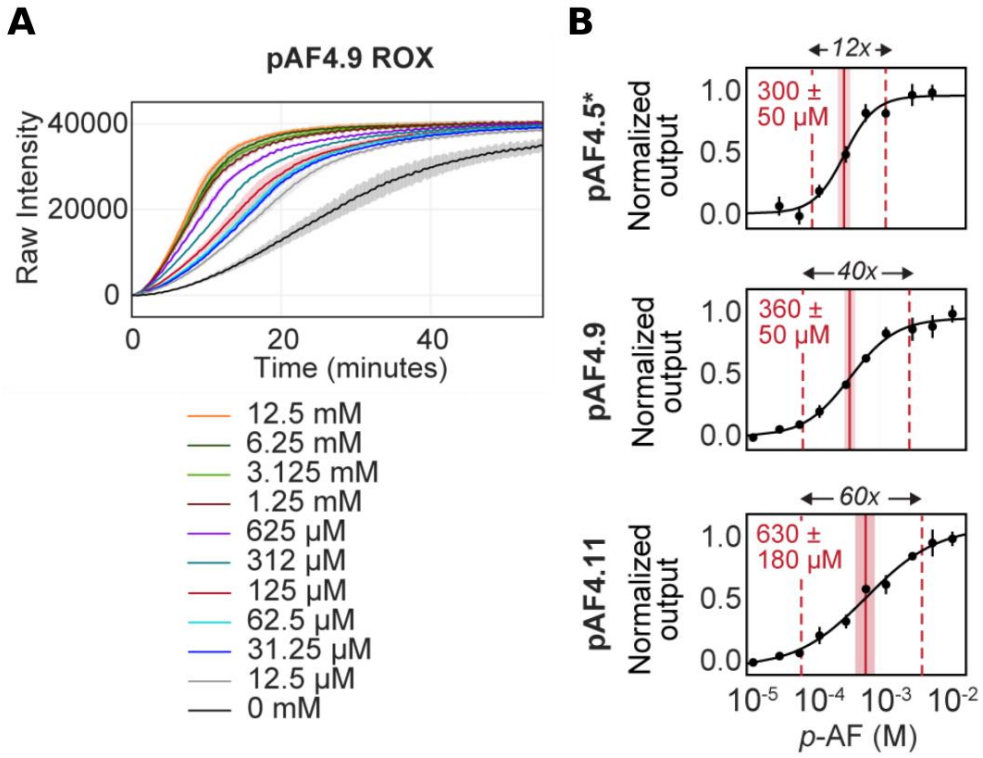


Figure 1.4. Responsiveness of designed *p*-AF kinetic ribosensors to titrated *p*-AF. A. Co-transcriptional traces of pAF4.9 ribosensor responses for a titration from 12.5 µM to 12.5 mM of *p*-AF. Clear, concentration-dependent behavior that can be used to quantify titers can be derived from the linear responses in the first 10 minutes of the assay. B. EC_{50}s for three *p*-AF responsive kinetic devices were calculated after fitting Hill functions to *p*-AF titration data, with the useful dynamic range shown between the dotted lines. Adapted with permission from Burke (2017).

1.6 Organization of thesis

This dissertation is laid out in four chapters, consisting of an introduction and then three chapters each detailing a distinct focus of my research. I started research in the Carothers Lab in January 2014; this thesis covers work carried out between the summer of 2015 and the summer of 2019.

The first chapter outlines both my motivation for pursuing a PhD in bioengineering focusing on metabolic engineering and covers introductory material on topics relevant for my thesis for an audience unfamiliar with the field. This includes a thorough overview of prior research in the Carothers Research Group covering kinetic aptamer ribosensors, to distinguish my contributions from existing work that had been done on this platform.

The second chapter, "*In vitro* selection of an RNA aptamer against *p*-aminocinnamic acid," details both work done on establishing a label-free affinity chromatography approach for *in vitro* selections and the implementation of that method in selecting for an RNA aptamer binding *p*-ACA. Part of this work was published in an article entitled "Label-free selection of RNA aptamers for metabolic engineering" in the journal *Methods* in 2016. The results of the selection are part of a manuscript in preparation for submission to *ACS Synthetic Biology* entitled "Predicting RNA aptamer-small molecule binding affinities for applications in synthetic biology and metabolic engineering." This chapter is partially adapted from the text of those two publications.

The third chapter, "Predicting RNA aptamer-small molecule binding affinities for applications in synthetic biology and metabolic engineering," describes work on mining the dataset of existing RNA aptamers to select chemical characteristics of small molecules that best

predict RNA aptamer binding affinities. This work is in preparation for submission alongside the results of the *p*-ACA aptamer selection in a manuscript bearing the same title as this chapter.

The fourth chapter, "Probing properties of kinetic RNA aptamer ribosensors for biosensing in metabolic engineering," explores my contributions to the development of the kinetic aptamer ribosensor platform that uses RNA transcription and folding trajectories to control the sensitivity of a fluorescent output to the presence of a small molecule target. This work is included in an article entitled "Multi-state design of kinetically-controlled RNA aptamer ribosensors" and is in revision after submission to *Nature Chemical Biology*.

Chapter 2. *In vitro* selection of an RNA aptamer against *p*-aminocinnamic acid

Abstract

In metabolic engineering, RNA aptamer-based biosensor devices provide the means to implement dynamic, metabolite-responsive genetic control and screen for metabolite production in a quantitative, high-throughput manner. We select an aptamer for *p*-aminocinnamic acid, a precursor to synthesizing cinnamic acids, to assist in pathway engineering for biosynthesis of *p*-aminostyrene in *E. coli*. We detail the methods of a label-free affinity chromatography approach that allows for tracking of RNA across the course of an *in vitro* selection without radiolabeling or specialized equipment such as SPR. An aptamer with a dissociation constant of 29 μM was selected using this label-free affinity chromatography approach but reported slow binding kinetics. We discuss the challenges of applying small molecule aptamers to metabolic engineering work, highlighting the role played by aptamer dissociation constants, intermediate and product titers, biosensor architecture, and association kinetics.

2.1 Introduction

Biosensors play an important role in metabolic engineering as means for dynamic regulation of gene expression and for metabolite screens. The versatility of RNA aptamer-based devices for implementing these aims, the ease of generating RNA aptamers relative to generating protein binders, and the wealth of secondary structure prediction tools that allow us to incorporate aptamers into RNA devices make aptamers a promising biosensor for use in metabolic engineering. *In vitro* selections can yield RNA aptamers through the iteratively screening of a diverse nucleic acid pool against an immobilized ligand.

The progress and success of an aptamer selection is monitored by quantifying the relative amount of RNA that is eluted with the ligand after each round of selection. Many current approaches for nucleic acid quantification require radioactive labeling[21]; label-free methods, such as monitoring the progress of a selection by surface plasmon resonance (SPR), often require expensive or specialized equipment[47]. Although these methods have high sensitivity for tracking enrichment for binders, they are impractical for many researchers due to cost or lack of access. While not as sensitive as radioactive labelling or SPR, quantifying nucleic acid using a small volume spectrophotometer such as a Thermo Scientific Nanodrop can still accurately track the progress of an RNA aptamer selection if sufficient quantities of RNA are transcribed and applied to the ligand-immobilized column.

Affinity chromatography also holds advantages over newer developments such as CE SELEX or microfluidic SELEX (M SELEX). Large pool sizes of 10^{12} to 10^{15} are necessary for the selection of high affinity aptamers for metabolic engineering. The informational complexity of high affinity aptamers for small molecules makes the expected number of such aptamers

extremely low even in pools of over 10^{12} sequences[48]. Volume limitations make M SELEX and CE SELEX impractical for selecting from pools of larger than about 10^{13} sequences[30,33]; affinity chromatography on columns does not face a similar limitation, making it more suitable for selection of high affinity aptamers. Affinity chromatography on a column also provides a straightforward method of generating specific aptamers through counterselection. High affinity aptamers are not necessarily highly specific aptamers, so counterselection using chemically similar compounds that occur physiologically can screen out non-specific binders[48]. Affinity chromatography also allows for off-rate selection in physiological buffer conditions.

The immobilization chemistry described here has been applied for immobilizing fluorophores such as tetramethylrhodamine (TMR) and amino acid derivatives such as *p*-aminophenylalanine (*p*-AF) and *p*-aminocinnamic acid (*p*-ACA) to an amino polyethylene glycol polyacrylamide copolymer (PEGA) resin for affinity chromatography. It could be further applied for immobilizing other amino acid derivatives, fluorophores, or aromatics, including compounds relevant to recent engineered metabolic pathways[26,49]. The immobilization chemistry described derives from the use of Fmoc-protected amino acids for solid phase peptide synthesis[50,51] and can be implemented by most labs that conduct chemical or biological work.

We demonstrate this approach by selecting an aptamer for *p*-ACA, a metabolic intermediate in the *p*-AS engineered biosynthetic pathway. Through analysis of the resultant aptamer and its mismatched binding affinity with its intended application, we draw important lessons on matching the sensitivity of an aptamer device to its application, motivating the work of later chapters.

2.2 Methods

2.2.1 Ligand immobilization onto PEGA column

Ligand immobilization of *p*-ACA was achieved through three steps: coupling of the ligand, acetylation of unreacted amines, and Fmoc deprotection (Fig. 2.1). Initially, Fmoc protected *p*-ACA was coupled onto an amino PEGA resin. The unreacted amines on the resin were then acetylated with acetic anhydride to prevent RNA from sticking to the unreacted amines. After the unreacted amines were capped, the Fmoc group protecting the amines on *p*-ACA was deprotected, leaving *p*-ACA immobilized on the amino PEGA resin with all other amines on the resin acetylated and capped.

2.2.1.1 Coupling

200 μL selection columns were made by coupling Fmoc-*p*-aminocinnamic acid (Fmoc-*p*-ACA) to an amino PEGA matrix. 0.11 g amino PEGA beads (Novabiochem 855015) were swelled in 40 column volumes (CVs) water in Bio-Rad Poly-Prep Chromatography Columns (Bio-Rad 7311550), where a column volume was defined as 200 μL. Amino PEGA was then washed with 40 CVs dimethylformamide (DMF). 50 μL of 12 mM Fmoc-*p*-ACA in DMF, 100 μL of 12 mM diisopropylethylamine (DIPEA) in DMF, and 50 μL of 12 mM PyBOP in DMF, were added to the 200 μL of amino PEGA resin for coupling. The column was covered with foil to decrease exposure to light and was shaken on a small rotator at 30 rpm at room temperature overnight (12-18 hours). In this reaction, Fmoc-*p*-ACA couples with the amino PEGA resin by the forming a peptide bond.

2.2.1.2 Acetylation

At the end of overnight coupling, the column was washed with 40 CVs of DMF by filling the column with 8 mL DMF, capping it, agitating the column, and then draining the DMF after the beads settled. The column was then acetylated with a mix of 16.8 µL acetic anhydride, 31.2 µL DIPEA, and 152 µL DMF. The column was covered with foil and the acetylation was shaken at room temperature for 2 hours. Acetylation caps the amine ends of the amino PEGA resin and prevents aptamers that can interact with the amine end of *p*-ACA from interacting with the amines on the PEGA resin. At the end of acetylation, the column was washed with 40 CVs of DMF.

2.2.1.3 Deprotection

Fmoc deprotection was performed by shaking the column with 1 mL 20% piperidine in DMF for 30 minutes. The absorbance of the deprotection flow through at 301 nm was measured to determine the amount of Fmoc released from the column and thus the extent of coupling; most columns exhibited a final concentration of 0.5 mM *p*-ACA immobilized to the amino PEGA resin. Fmoc deprotection removed the protecting groups from the amines on *p*-ACA, leaving just *p*-ACA conjugated to the amino PEGA resin by a peptide bond.

2.2.1.4 Column preparation for affinity chromatography

The column was washed again with 40 CVs of DMF, then soaked with 40 CVs of water and 20 CVs of selection buffer (SB) for half an hour each. The water and selection buffer washes were important due to the different swelling properties of the PEGA resin in DMF and water. Selection buffer was made with 130 mM K glutamate (pH 7.5), 15 mM NaCl, 10 mM

dithiothreitol (DTT), and 1 mM $MgCl_2$. After incubation with SB, the column was ready for affinity chromatography.

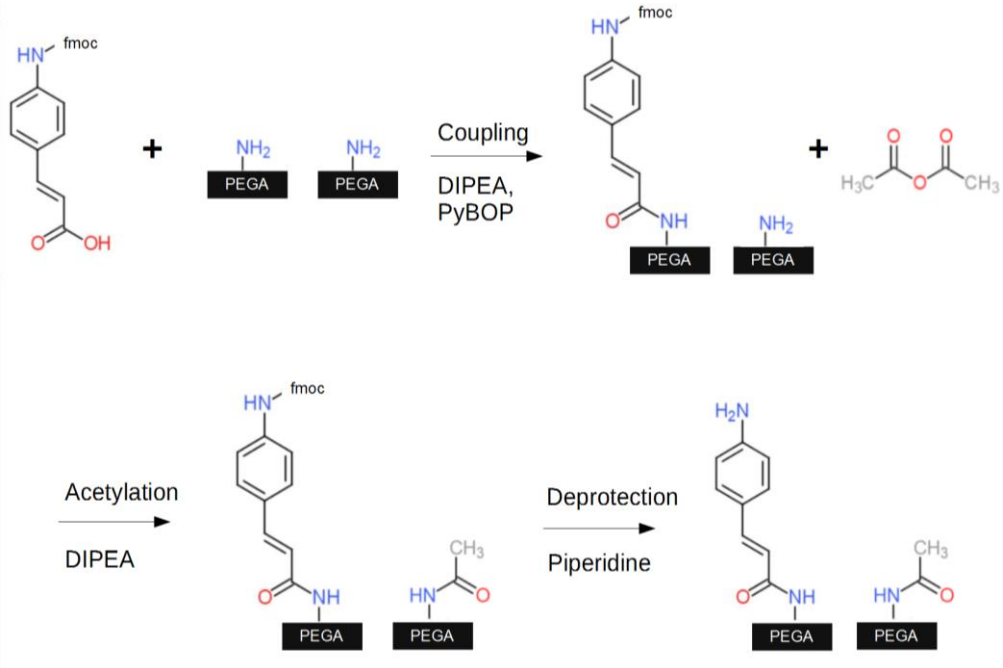


Figure 2.1. Simplified reaction scheme for the immobilization of *p*-ACA onto amino PEGA resin using Fmoc. In the coupling reaction, DIPEA and PyBOP are used to form a peptide bond between Fmoc-*p*-ACA and the amino PEGA resin. Following coupling, some unreacted amine groups are left on the PEGA resin, which must be capped to prevent RNA from aggregating on the polar amines. Acetic anhydride and DIPEA are used to acetylate unreacted amines. All unreacted amines on PEGA resin are acetylated, while amines on *p*-ACA are protected by Fmoc. Immobilized *p*-ACA deprotected of Fmoc using piperidine. After deprotection, *p*-ACA is immobilized on amino PEGA resin and unreacted amines are capped and the ligand is prepared for affinity chromatography. All steps were performed at room temperature.

2.2.2 Preparation of DNA pool

A library of 10^{14} DNA oligonucleotide sequences was generated using a previously described pool design strategy from Integrated DNA technologies[16]. The pool included a 12-base

stem tetraloop flanked by two randomized regions of 28 bases each, which in turn were flanked by constant regions that were used for PCR amplification of the pool. The pool was resuspended in water and size purified using an 8% PAGE gel and a Whatman Elutrap (GE Healthcare). A 1 μM primer extension with a primer carrying the T7 promoter was used to convert the oligonucleotides into transcription-ready dsDNA.

2.2.3 *In vitro* transcription and preparation of RNA

In each round, candidate RNA sequences were produced from the DNA pool in 0.5 μM *in vitro* transcriptions. Transcriptions were performed in 25 mM $MgCl_2$, 10 mM DTT, 2.5 mM spermidine, 40 mM Tris (pH 7.8), and 0.01% Triton X-100. 9 mM GTP and 5 mM each of the other rNTPs were added along with 1 U/mL Thermostable Inorganic Pyrophosphatase (NEB M0296), 2.5 U/μL T7 RNA Polymerase (NEB M0251), and 1 U/μL RiboLock RNase inhibitor (ThermoFisher EO0381). Transcription of the DNA pool in the first round was determined by the size of the DNA pool; a nanomole pool (10^{14} sequences) would require a 2 mL transcription volume and transcribing a larger pool requires scaling accordingly. Subsequent transcriptions were performed at 200 μL volumes to produce enough RNA for label-free detection. Transcription buffer should be made fresh to ensure high transcription yields; all buffer components should be stored at -20 ºC prior to use.

Transcriptions were run overnight (12-18 hours) at 37 ºC. Transcriptions were terminated by adding DNaseI (NEB M0303), 0.5 U/μg DNA, for 30 minutes at 37 ºC. At the end of DNaseI incubation, 25 mM EDTA was added and DNaseI was heat inactivated at 75 ºC for 10 minutes. 200 μL transcriptions were then desalted using NAP-5 Sephadex columns (GE Healthcare 17085302) and RNA was eluted in water. RNA in water was ethanol precipitated with KCl and

GlycoBlue co-precipitant (ThermoFisher AM9516), then resuspended a smaller volume of water to be run out on a PAGE gel (typically about 100-200 μL).

The resuspended RNA was run on a 1.5 mm PAGE gel at 13 W for 1.5 hours; the correct bands were identified using UV shadowing by comparison with xylene cyanol and bromophenol blue dye bands and then cut from the gel. Typically, a 100-base RNA migrates slightly slower than the xylene cyanol band on an 8% PAGE gel; RNA size can also be determined by comparison with a RNA ladder. RNA was recovered from the cut gel slices with 0.5X TBE using a Whatman Elutrap; for recovery, the Elutrap was operated at 200 volts for 2 hours. Eluted RNA was extracted from the Elutrap, ethanol precipitated with KCl, and then resuspended in water. Resuspended electroeluted RNA totaled between 60-200 μg of RNA. The total mass of RNA must not be lower than this range to ensure reliable quantification of eluted RNA from the column after affinity chromatography.

2.2.4 *In vitro* selection

Ethanol precipitated RNA was resuspended in 200 μL water and then incubated at 37 ºC for 5 minutes; 4 μL 200 μM yeast tRNA, 4 μL 1 mg/mL BSA, and 200 μL of 2X SB were then added. The RNA mixture was then applied to the ligand-immobilized column and incubated for 5-30 minutes. Columns were washed with 20 CVs SB, then eluted with 2 CVs 1 mM *p*-ACA (Pfaltz and Bauer, 95% purity) in SB three times. Another 2 CVs 1 mM *p*-ACA in SB were applied to the column and collected 1 hour later. All selection steps were performed at room temperature unless otherwise noted. Off-rate selection can be performed by varying the amount of time the ligand in SB is incubated on the column.

2.2.5 Reverse transcription and pool amplification

Elution fractions were pooled and desalted in NAP-25 Sephadex columns (GE Healthcare 17085201) and eluted in water. Desalted RNA was volume reduced by butanol extraction and then cleaned up again by ethanol precipitation with KCl. RNA was resuspended in 50 µL water and the nucleic acid concentration was measured by Nanodrop. RNA was reverse transcribed using illustra Ready-to-Go RT-PCR beads (GE Healthcare 27925901) at 42 ºC for 1 hour. RNA was mixed with RT primer and heated to 75 ºC for 5 min, then chilled on ice prior to being added to the RT-PCR reaction.

RT reactions were then pooled and cleaned up using the Qiagen PCR Purification Kit (Qiagen 28104). Purified cDNA was then PCR amplified using DreamTaq DNA Polymerase (ThermoFisher K1081) to a total of 10 µg DNA. The Qiagen PCR Purification Kit was used again to clean up the PCR-amplified DNA, which was then used for transcription of the next round of the selection. It was essential to monitor the progress of PCR reactions when amplifying the pool: overamplification of the pool allows for mispriming and the amplification of incorrect products. We monitored amplification by performing Taq PCR incrementally and using agarose gels to qualitatively assess PCR progress, although monitoring using qPCR should be effective as well. After each round, half of the amplified DNA was used for transcription in the next round while half was stored at -20 ºC for later sequencing and as a backup in case of unexpected failure in a subsequent round.

After a sustained increase in the percentage of RNA recovered in the *p*-ACA elutions was observed, PCR products were cloned using TA cloning and sequenced; alternatively, the PCR amplified DNA can be assessed with the sequencing technology of your choice.

2.2.6 RNA quantification and label-free tracking

During the selection, we quantified nucleic acid amount at three points in each round:

1. DNA pool quantification prior to *in vitro* transcription
2. RNA pool quantification following the ethanol precipitation after electroelution
3. Quantification of pooled ligand-eluted RNA from affinity chromatography column following ethanol precipitation

Quantification of the DNA pool was performed at the beginning of each round to determine the volume of DNA to load in 0.5 μM *in vitro* transcription reactions. Typically, 100 μmoles of DNA were loaded into each 200 μL transcription. Post-electroelution RNA pool quantification was necessary to determine the amount of correct-size transcription product and thus the amount of RNA being applied to the column. After Round 1, most 200 μL transcription yielded between 60-200 μg of the correct length product.

Monitoring ligand-eluted RNA allowed us to track the progress of the selection. In most early rounds, less than 1% of the RNA applied to the column would be eluted with the ligand. GlycoBlue co-precipitant used in ethanol precipitation contributes to nucleic acid absorbance measurements on a small volume spectrophotometer and can cause overestimation of RNA amount, especially when the total eluted RNA amount is less than 500 ng. While this problem can be circumvented by ensuring that sufficient RNA (greater than 60 μg) is applied to the column each round, the lower sensitivity of tracking using spectrophotometry is not problematic for tracking the selection as a whole: once enrichment was measurable and the binders form a large portion of the pool, RNA recovered from the column reached higher than 20%. Even if only 60 μg of RNA were applied to the column, 20% recovery would correspond to 12 μg RNA,

which is easily detectable on a Nanodrop. It is impractical to perform the selection with lower amounts of RNA, as using less RNA may cause the concentration of resuspended ligand-eluted RNA to fall below the limit of detection for the Nanodrop, which is between 2-3 ng/μL of nucleic acid for most Nanodrop models.

2.2.7 Reselections

For reselections, doped reselection pools were synthesized at the Keck facility to achieve percentages of 79% of the initial base and 7% of each other base at every doped position in each of the three reselection pools. 1×10^{14} DNA sequences from each pool were transcribed for the initial round of the reselection; each round of the reselection was run with the same approach as the initial selection, except that no counterselection was applied.

2.2.8 Sequencing

Linear pool DNA was cloned into plasmid vectors via TA cloning using the Thermo Scientific InsTAclone PCR Cloning kit (#K1214) and chemically transformed into competent DH10B *E. coli*, which were plated on LB agar plates containing 50 ug/mL carbenicillin with 40 uL 20 mg/mL X-gal in DMF and 40 uL 100 mM IPTG coated on each plate. Blue white screening was employed to pick colonies containing inserts; 20 colonies were picked for overnight cultures in LB and then mini-prepped using a Qiagen Kit and then sent for sequencing. Resultant sequences were aligned in Jalview using the ClustalO algorithm; phylogenetic tree analysis was used to sort similar sequences into families. Representative sequences from each family were synthesized, transcribed, and tested for binding against the target ligand using affinity chromatography binding assays.

2.2.9 Affinity chromatography binding assays

Affinity chromatography binding assays were performed to determine the percent binding of particular sequences to *p*-ACA-functionalized affinity chromatography columns. In brief, assays were performed under the same *in vitro* transcription, ligand immobilization, and affinity chromatography conditions as the *in vitro* selection itself; percent recovery of RNA applied to the columns was quantified after elution of binders from the column with *p*-ACA in selection buffer, desalting, and volume reduction.

2.2.10 Biolayer interferometry characterization of aptamer kinetics

The binding affinity and kinetics of the selected *p*-ACA aptamer were characterized using biolayer interferometry; DNA aptamer binding kinetics for a small molecule have been previously demonstrated[52]. Aptamer DNA templates were built with a T7 promoter and a 3' tail designed to minimize interactions with the aptamer sequence. DNA templates were transcribed in the conditions indicated above to generate RNA aptamers with 3' tails, which were then annealed with 5' biotinylated DNA adapters (Supplementary Fig. S2.1). RNA aptamers were combined at a 2:1 ratio with DNA adapters in 10 mM Tris, 50 mM NaCl, and 1 mM EDTA. Samples were heated to 85 ºC for 1 minute and cooled -1 ºC every 30 seconds until reaching 25 ºC. Annealing reaction was cleaned up using a Qiagen PCR clean-up kit and diluted into selection buffer.

Biolayer interferometry was run using a Pall Fortebio Octet Red96 with Super Streptavidin biosensors. Sensors were equilibrated in selection buffer for 10 minutes before the assay and then for 2 minutes on the Octet before the annealed RNA aptamer with biotinylated

DNA adapter was loaded for 3-10 minutes, dependent on loading efficiency. On and off rate kinetics and dissociation constants were calculated using a monophasic fit of both association and dissociation using Fortebio Octet Data Analysis software.

2.2.11. Safety

The immobilization procedure, selection steps, PAGE gel casting, and butanol extraction should be performed in a chemical fume hood with standard personal protective equipment and face and eye protection. Hazardous waste from organic solvent washes and waste from butanol extraction should be collected separately and disposed of in accordance with institutional regulations.

2.3 Results and Discussion

2.3.1 Label-free approach for selecting an RNA aptamer

The progress and success of an aptamer selection is monitored by quantifying the relative amount of RNA that is eluted with the ligand after each round of selection. Many current approaches for nucleic acid quantification require radioactive labeling[16,21]; label-free methods, such as monitoring the progress of a selection by surface plasmon resonance (SPR), often require expensive or specialized equipment[36]. Although these methods have high sensitivity for tracking enrichment for binders, they are impractical for many researchers due to cost or lack of access. While not as sensitive as radioactive labeling or SPR, quantifying nucleic acid using a small volume spectrophotometer such as a Thermo Scientific Nanodrop can still accurately track the

progress of an RNA aptamer selection if sufficient quantities of RNA are transcribed and applied to the ligand-immobilized column.

To compare this simple label-free approach to previously reported radiolabeled tracking approaches for *in vitro* selection, affinity chromatography assays for binding of the previously reported *p*-AF R1-1 and 4z1d3 aptamers were repeated using the label-free approach[16]. For both aptamers, the measured percent bound to the affinity chromatography columns were slightly lower but matched at least 80% of the radiolabeled levels of binding (Fig. 2.2). As the label-free approach requires additional RNA purification and clean-up steps before measurement that typically result in small losses of total RNA, the reported binding is consistent with expectations for a small volume spectrophotometer-based approach. Critically, this experiment demonstrates that enriched RNA recovery yields are quantifiable with a label-free approach.

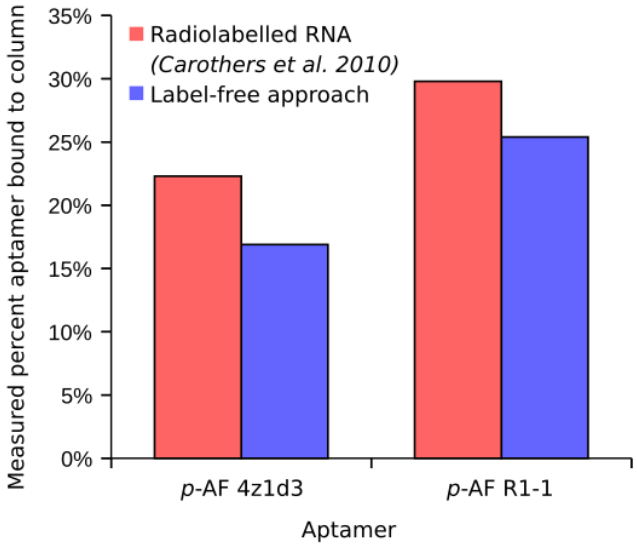


Figure 2.2. Comparison of measurements of aptamer bound to affinity chromatography columns for radiolabeled RNA and the presented label-free approach. Radiolabeled RNA percentage bound was reported in Carothers et al. 2010; although the label-free approach reported a lower

percentage bound than the radiolabeled approach for both the 4z1d3 and R1-1 *p*-AF aptamers, results were consistent with expected loss of RNA in purification steps.

2.3.2 *In vitro* selection for RNA aptamer against *p*-aminocinnamic acid

In vitro selection was initiated from a synthesized DNA pool with 2.8×10^{14} diversity; the pool consisted of 102-base sequences with two 28-base randomized regions sandwiching a 12-base structured stem tetraloop, bookended with two constant regions for amplification. The stem tetraloop was included in the pool design as previous selections have indicated that seeded structures may prepay a portion of the information content necessary to specify an aptamer[48].

During initial rounds of selection, we recorded less than 0.5% RNA recovery through elution from the column with *p*-ACA. After 12 rounds of selection, we observed 4% RNA recovery after affinity chromatography; the selection was ended at 14 rounds after the RNA recovery bump was sustained and DNA from the last three rounds were sequenced (Fig. 2.3a). Three sequences families (*p*-ACA_A, *p*-ACA_B, and *p*-ACA_C) with high similarity yielded multiple counts across multiple rounds; representative sequences from each of the three sequences families, complete with constant regions, were individually tested for binding using affinity chromatography and exhibited binding between 2.3% and 4.2%. After verifying binding of each sequence family, the minimum free energy structure of each sequence was analyzed using Vienna RNAfold[42] and large structured regions were identified in each family; sequences were truncated to contain only the sequence that would fold into the putative MFE structures.

2.3.3 Optimization and characterization of an RNA aptamer binding *p*-aminocinnamic acid

Three doped reselection pools were designed based on the three sequence classes yielded from the initial selection. Doped pools were generated with 79% fidelity to the starting

sequences that had been tested for binding, with 7% doping of each of the other bases. We performed five rounds of reselection on each of the three pools *in vitro*, stopping after RNA recovery exceeded 3% in two consecutive rounds (Fig. 2.3b). 38 members of the three resultant pools were sequenced. Analysis of conserved bases after sequencing revealed that the three reselection pools contained a single 38-base conserved aptamer sequence, with 18 bases fully conserved across all sequences obtained from the three pools (Fig. 2.3c). The putative structure for this aptamer, based on the minimal free energy fold, includes two bulges (Fig. 2.3d). Base pair variant analysis across the three sequence classes suggested that flanking stems of four base pairs and two base pairs were necessary on either side of the conserved aptamer core. The hairpin was displayed on different ends of the aptamer for *p*-ACA_C compared to *p*-ACA_A and *p*-ACA_B, demonstrating that the hairpin is not critical to the binding structure.

The most favored sequence suggested by sequence analysis exhibited 22% binding, a fivefold increase in column binding over the naïve aptamer sequence resultant from the initial *in vitro* selection (Fig. 2.3e). Against a negative control affinity chromatography column lacking immobilized *p*-ACA, the aptamer exhibited 0.6% binding, confirming its specificity to binding the target and not the column (Fig. 2.3e).

The aptamer was confirmed to bind to *p*-ACA in solution via bio-layer interferometry (BLI); through BLI kinetic analysis, we found a K_D of 29 ± 6 μM between the reported aptamer and *p*-ACA (Supplementary Fig. S2.2). The on-rate for the *p*-ACA aptamer was reported to be 37 ± 6 $M^{-1}s^{-1}$ through BLI analysis, which was over an order of magnitude slower than characterized rates for the *p*-AF 4z1d3 aptamer (490 $M^{-1}s^{-1}$).

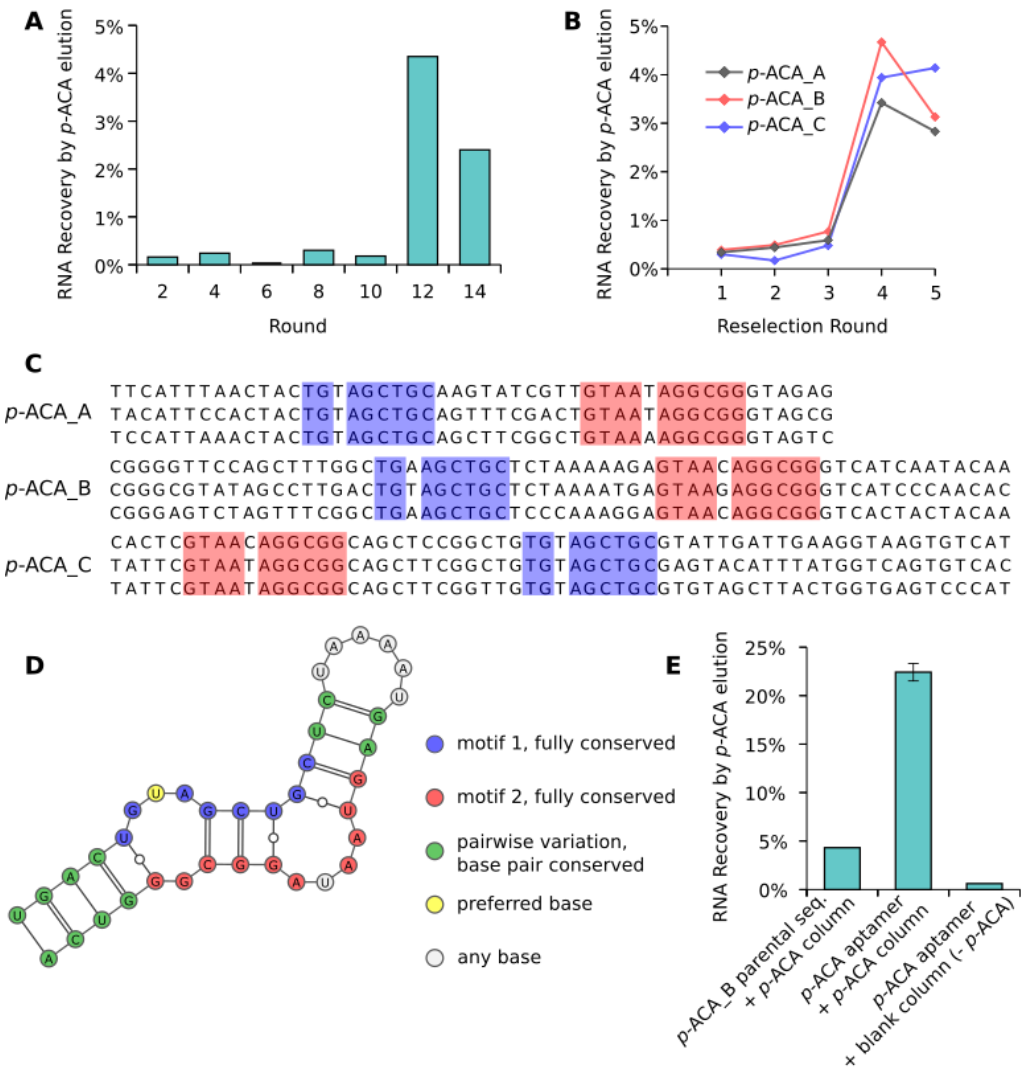
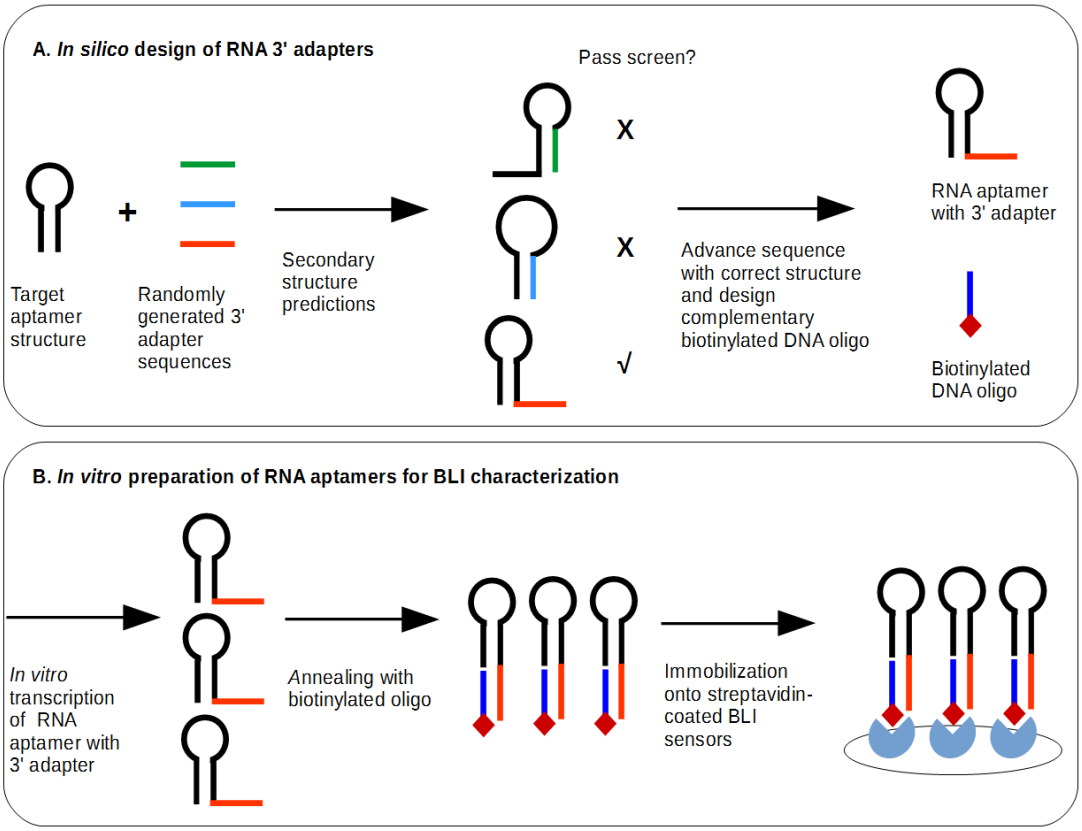


Figure 2.3. *In vitro* selection for an RNA aptamer to bind *p*-ACA. A. Enrichment of RNA binders was observed after 12 rounds of initial *in vitro* selection from a pool of 2.8 x 10^{14} diversity. The initial selection was run for 14 rounds total. B. Reselection with three doped pools resultant from sequences recovered in initial *in vitro* selection. All three reselection pools exhibited >3% binding after four rounds. C. Representative samples from sequencing results of enriched reselection pools. All three pools displayed two fully conserved motifs totaling 18 conserved bases. D. Putative secondary structure of the *p*-ACA aptamer, with bases colored to indicate degree of conservation, predicted by Vienna RNAfold. E. RNA recovered from affinity chromatography for *p*-ACA_B pool parental sequence, optimized *p*-ACA aptamer, and the optimized *p*-ACA aptamer screened against a column without *p*-ACA functionalized. Optimized aptamer exhibits higher binding than the parental sequence and the no *p*-ACA negative control.

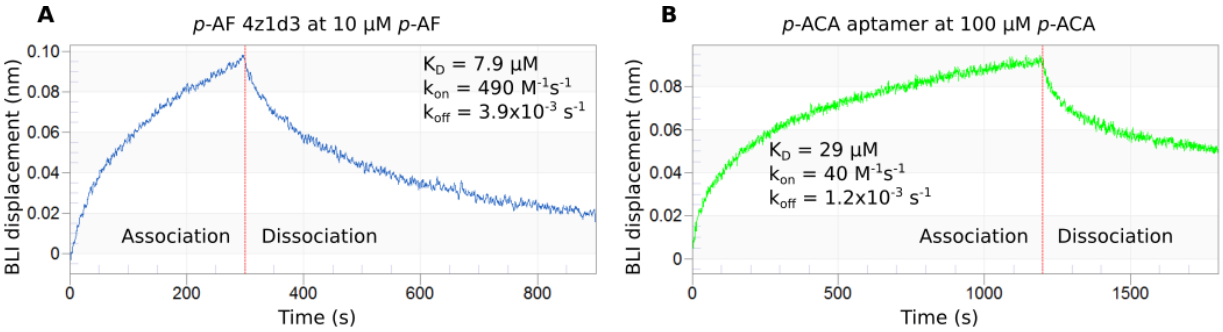
In current pathway engineering work in the Carothers Lab, *p*-AF titers of 200 mg/L, about 1 mM, have been achieved; *p*-ACA has been produced by *E. coli* expressing PAL2 from spiked *p*-AF; about 1% *p*-AF to *p*-ACA conversion has been reported. Two challenges have arisen that make useful *p*-ACA sensors currently unfeasible. First, the binding affinity of the *p*-ACA aptamer (~30 uM) is insufficient to interact with existing biosynthetic titers of the small molecule. Second, based on the association rate of the *p*-ACA aptamer and Eq. 1.1, the binding time window t_{bind} would have to be on the order of a minute to construct a kinetic aptamer ribosensor with an EC_{50} below 1 mM. That binding time that is not yet achievable with current approaches for tuning sensitivities for kinetic aptamer ribosensors; the longest predicted t_{bind} is for pAF4.5, which has a t_{bind} of about 4.5 seconds.

This highlights a key problem with utilizing biosensors of any kind: a biosensor will only have utility if, in the context of its application, it has binding properties that will allow it to function for its intended purpose. In other words, merely having an aptamer for a metabolite and being able to produce some level of that metabolite is insufficient to give utility to an aptamer if the sensitivity of that aptamer is not matched to concentration of metabolite that needs to be sensed. Yet another dilemma is presented here: at present, only some 60 small molecule RNA aptamers have been reported; if we try to use aptamers in metabolic engineering only in situations where we already know they will work, we will find ourselves boxed in to a very limited set of applications. Yet the alternative option is to start from scratch, committing to a time-consuming aptamer selection without any knowledge that the resulting aptamer would be useful to sense the target small molecule at the desired concentrations. Therefore, there is a need for approaches for *a priori* predictions on whether RNA aptamers may be useful for biosensing for specific metabolic engineering applications.

2.4 Supplementary Materials



Supplementary Figure S2.1. Setup for immobilizing RNA aptamers onto streptavidin-coated biosensors for BLI. A. *In silico* design was used to design RNA 3' adapters that would allow an aptamer to hybridize with a biotinylated DNA oligo while maintaining the appropriate secondary structure. B. Designed RNAs with 3' adapters were transcribed *in vitro* from DNA templates and annealed with complementary biotinylated oligos, which were then immobilized onto Fortebio Super Streptavidin biosensors via biotin-streptavidin interactions for BLI analysis on the Octet.



Supplementary Figure S2.2. BLI characterization of kinetics for the (A) *p*-AF and (B) *p*-ACA aptamers. *p*-AF 4z1d3 aptamer kinetics were measured at 10 µM *p*-AF, with a reported k_{on} of 490 $M^{-1}s^{-1}$ and a K_D of 7.9 µM. The *p*-ACA aptamer was assayed at 100 µM *p*-ACA, with a K_D of 29 µM and a k_{on} of 40 $M^{-1}s^{-1}$.

Supplementary Table S2.1. *p*-ACA aptamer selection sequences. Doped sequences (79% indicated base, doped with 7% each of the three other bases) are underlined for the reselection pools.

Name	Sequence
p-ACA aptamer selection naïve pool	GAGGTGTAGTGATGACNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGCCGAAGC TGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAGATGCGCGATCC
Pool reverse primer	CTCCGAGGTGTAGTGATGAC
Pool Extension primer	GCTAATACGACTCACTATAGGGCTGGATCGCGCATCTCG
p-ACA_A	CTCGGGGGTTCGCATTAAACTACTGTAGCTGCAGCTTCGGCTGTAAAAGGCGGGTAGA GTGCTTAAGGTGGGTCATCACTACACCTCGGAGCCC
p-ACA_B	CCGGGGGTCTAGCTTCGGCTGTAGCTGCTCTAAAATGAGTAACAGGCGGGTCATCACTA CACCTCGG
p-ACA_C	AGGTGTTATTCGTAATAGGCGGCAGCTTCGGCTGTGTAGCTGCGTGTAGATTAATGGTC AGTGTCATCACTACACCT
p-ACA_A reselection pool	GTCGGACCACTAAGGTGTGCACTCTACCCGCCTTTTACAGCCGAAGCTGCAGCTACAGT AGTTTAATGCACACCTGGAGGTCCCTA
p-ACA_B reselection pool	GTCGGACCACTAAGGTGTCGAGGTGTAGTGATGACCCGCCTGTTACTCATTTTAGAGCA GCTACAGCCGAAGCTAGACCCCCGACACCTGGAGGTCCCTA
p-ACA_C reselection pool	GTCGGACCACTAAGGTGTAGTGATGACACTGACCATTAATCTACACGCAGCTACACAGC CGAAGCTGCCGCCTATTACGAATAACACCTGGAGGTCCCTA
Reselection pool reverse primer	GTCGGACCACTAAGGTGT

Reselection pool extension primer	GCTAATACGACTCACTATAGGGTAGGGACCTCCAGGTGT
p-ACA aptamer w/ T7 promoter	GCTAATACGACTCACTATAGGGTGGCTGTAGCTGCTCTAAAATGAGTAATAGGCGGGTCACCC
p-ACA aptamer with BLI adapter tail	GGGTGGCTGTAGCTGCTCTAAAATGAGTAATAGGCGGGTCACCCGAAGTAGTGTTGAGAAGCCTCAGT
Biotinylated oligo adapter for BLI	/biosg/ ACTGAGGCTTCTCAACACTACTTC

Chapter 3. Predicting RNA aptamer-small molecule binding affinities for applications in synthetic biology and metabolic engineering

Abstract

RNA aptamers are increasingly utilized for building biosensors and genetic regulators for metabolic engineering, with roles as diverse as sensing metabolic intermediates to effect pathway control and *in vitro* biosensing to assess yields of different engineered strains. *In vitro* selection provides a straightforward method of generating RNA aptamers to bind metabolic intermediates that currently lack known sensors and regulators, but uncertainty in binding affinities of aptamers resultant from selections hinder the broader use of aptamers as bioreceptors for small molecule targets. Here, we present what is, to our knowledge, the first method for *a priori* estimation of RNA aptamer binding affinities against small molecule targets. We apply a multiple linear regression approach to construct a model for estimating aptamer binding affinities from the chemical characteristics of 31 small molecules with existing RNA aptamers. We demonstrate that this model can generate accurate predictions for multiple small molecules, and we generated predicted ranges of binding affinities for several metabolites relevant to the metabolic engineering community. We demonstrate how these binding affinity predictions can be used in conjunction with reported metabolite titers to predict the utility of an RNA aptamer for sensing a given molecule and we discuss how to apply this to selecting pathways and molecules for RNA aptamer biosensing. This presented method provides the first approach to quantitatively predict binding affinities of small molecules for RNA aptamers.

3.1 Introduction

Small-molecule binders play critical roles in metabolic engineering as means for effecting metabolite-responsive changes in gene expression and for detecting and quantifying metabolite production in engineered cells[14,18]. However, many intermediates and products that are targets for metabolic engineering lack known sensors and regulators. RNA aptamers are a binding solution well-suited for many applications, as they are comparably easier to generate than protein binders for molecules lacking known sensors and they can be incorporated into genetic control and biosensing readout devices through design methods utilizing RNA secondary structure prediction tools[42,53]. Aptamers have been incorporated into fluorogenic biosensors for enzyme evolution[54], assembled into aptazymes for genetic control[14,17], assembled into riboswitches for translational control[43], and encapsulated in droplets to provide single-cell readouts for high-throughput screening of secreted metabolites[55], demonstrating the wide utility of these RNA structures as sensors in metabolic engineering (Fig. 3.1a).

Due to the difficulty of matching the binding affinity of RNA aptamers to the sensitivities required of their intended applications, the total number of reported RNA aptamers is quite small[27]. The time-consuming nature and uncertain outcomes of *in vitro* selections likely discourage the broader adoption of RNA aptamer devices as biosensors and dynamic genetic regulators in metabolic engineering. Anecdotally, as many as half of all *in vitro* aptamer selections fail to produce binders at all; in many other cases, such as the one detailed in this thesis, resultant aptamers have binding characteristics insufficient for the intended application[27]. This is a problem because small molecule aptamer binding affinities have been characterized to span seven orders of magnitude, from nanomolar binders to binders with K_Ds around 10 mM

(Fig. 3.1c). *In vitro* selections are neither temporally nor financially trivial, so any *a priori* information on a selection's potential would be valuable for deciding for or against utilizing aptamer biosensors.

Can we use our knowledge of existing small molecule RNA aptamers to predict the success and likely binding affinities that would result from *in vitro* RNA aptamer selections for other small molecules (Fig. 3.1b)? This question has been raised in the literature but so far there has been no serious attempt to quantitatively predict the binding affinities accessible for RNA aptamers to small molecules based on the properties of those small molecules. Extant work has explored the relationship between the informational complexity (rarity) of RNA aptamers and the binding affinity of those aptamers and found that lower informational complexity is needed to specify high affinity binders for small molecules with a higher molecular weight[16]. Here, we go further by developing a statistical model to look at the chemical factors that contribute to small molecule interactions with RNA aptamers. This statistical model can underlie workflows for aptamer sensor engineering that incorporate *a priori* predictions to inform decisions on aptamer *in vitro* selections (Fig. 3.1d).

Group contribution methods predict properties of small molecules by estimating a given property as the sum of contributions towards that property from the constituent functional groups of the molecule. Group contribution methods have been used to predict a wide range of thermodynamic properties of small molecules, from boiling and melting points of compounds to the binding affinities between drugs and their protein receptors[56,57]. Early group contribution methods considered free energy contributions exclusively from functional groups, relating a given property to a free energy calculation and determining the free energy relating to a given property for a molecule from the form:

$$\Delta G_{property} = \sum n_i \Delta G_i \qquad \text{Eq. 3.1}$$

where G_i is the free energy contribution towards a property for a functional group type *i* and n_i is the total tally of functional group type *i* in the molecule of interest.

By presenting methods for calculating the critical temperature, critical pressure, and critical volume of small molecules as a function of the sum of free energy contributions from the constituent functional groups of the molecules, Lydersen developed the initial framework for group contribution methods[58]. Joback and Reid expanded on Lydersen's work by developing group contribution predictions for properties such as a boiling point, heat of fusion, heat of formation, and liquid dynamic viscosity[59]. Andrews extended this method to drug-receptor interactions, modeling the free energy of binding with protein receptors as a property of small molecule drugs[56]. More recent work from Mavrovouniotis and Jankowski introduce interaction factors in their methods to account for free energy contributions from gross molecular structure[60,61]. Many group contribution methods- including all the examples cited here- are developed by training multiple linear regression models on databases of molecules and their properties.

Additive group contribution approaches are frequently utilized; previous reports have used multiple linear regression to calculate functional group weights for estimating standard Gibbs free energy of formation and drug-receptor binding energies[56,61]. In most group contribution methods, the property being estimated is related directly to some measure of free energy; the free energy term, in turn, is typically calculated as a sum of free energy contribution from constituent functional groups and from interaction effects. We reasoned that since protein-small molecule interactions have been successfully modeled using group contribution methods, nucleic acid-small molecule interactions could be amenable to this modeling approach as well.

Here, we present a modeling approach for predicting the binding affinity of small molecules to RNA aptamers based on the principles of group contribution methods. Based on the characterized binding affinities of 31 existing small molecule RNA aptamers, we applied multiple linear regression to estimate the contributions of chemical characteristics of small molecules to the binding energy between those small molecules and their cognate aptamers. We demonstrate that this model can generate accurate predictions for several small molecules and using this model, we generated predicted ranges of binding affinities for a number of metabolites relevant to the metabolic engineering community.

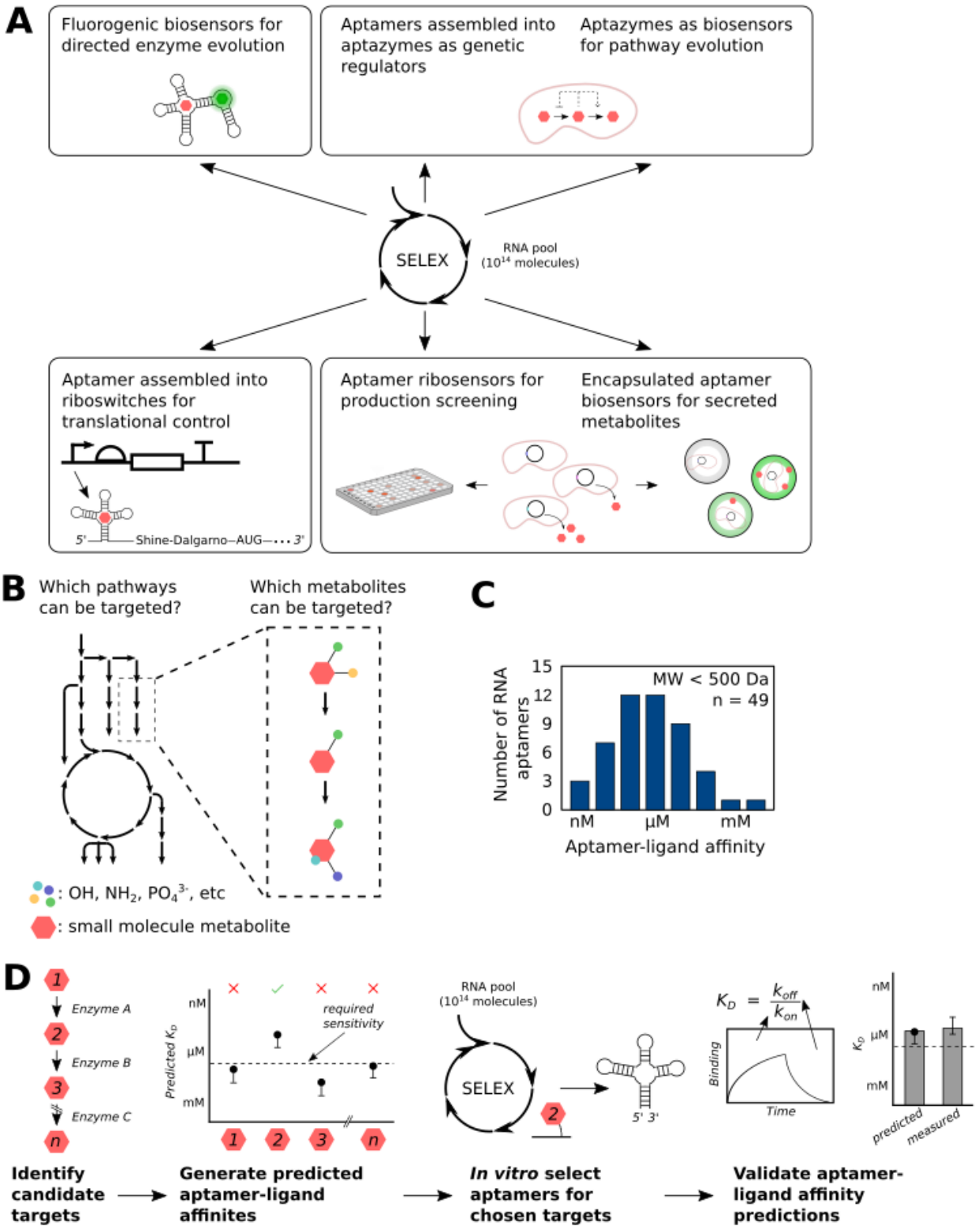


Figure 3.1. Although applicable in many biosensing applications, RNA aptamers remain difficult to use because of wide variances in binding affinities, so approaches to identify

pathways and molecular targets that may be more amenable to RNA aptamer-based devices are useful. A. *In vitro* selections have generated RNA aptamers that have been incorporated into a wide array of biosensing applications in metabolic engineering. B. Which metabolic pathways would benefit from RNA aptamer-based biosensing? What metabolic intermediates make the best targets for RNA aptamer devices? Can we answer these questions based on the chemical properties of candidate small molecule targets? C. Binding affinities for 49 small molecule RNA aptamers spans 7 orders of magnitude. This wide spread in binding affinities leaves large uncertainty in outcomes for new aptamer selections. D. A workflow for applying small molecule RNA aptamer binding affinity predictions for picking metabolites to target: metabolic intermediates and products can be assessed using the model to see if they match the required sensitivity for an application, then *in vitro* selections can be run for small molecules that meet the sensitivity criteria. Binding affinity characterization can then help validate the model prediction.

3.2 Methods

3.2.1 Modeling interactions between RNA aptamers and small molecules

Dissociation constants for RNA aptamer-small molecule binding can be interpreted as an equilibrium constant useful for determining the Gibbs free energy of reaction between the two species, where:

$$\Delta G_{bind} = -RTln(K_D) \qquad \text{Eq. 3.2}$$

Here, we assumed this binding energy to be primarily a property of the small molecule and that contributions to this binding energy correlated linearly with the chemical characteristics of small molecule targets. This allowed us to adapt a group contribution approach towards building a linear regression model to estimate the binding energies of small molecules for RNA aptamers in which contributions towards electrostatic interactions are assigned exclusively to the small molecule targets.

In order to construct a group contribution method, we made three major assumptions about the relationship between RNA aptamers and their small molecule targets. The first assumption was that the binding energy between an RNA aptamer and a small molecule was mostly dependent on the small molecule and was derived from the characteristics of a small molecule that would allow the molecule to form electrostatic interactions with RNA. Based on this assumption, we built a linear regression model that assigns the contributions towards these electrostatic interactions exclusively to the small molecule targets. Secondly, we assume that the dissociation constants reported for each aptamer of our data set reflect binding energies of the small molecules for RNA sequences that are likely to occur in RNA libraries of diversity less than 10^{15}. Thirdly, we assume an additive relationship between thermodynamic contributions of individual functional groups in a molecule and the overall binding energy of a molecule, an assumption that underlies most group contribution methods[56,58,59].

3.2.2 Compiling a dataset to train a model

To date, about 60 small molecules have characterized RNA aptamers generated by *in vitro* selection[27-29]. Reasoning that individual functional groups may no longer provide direct contributions to binding energy in larger small molecules, we culled all molecules greater than 500 Da from this aptamer data set. Remaining small molecules in the set were then evaluated for outlier functional groups; molecules with unique functional groups were removed as well. This left a dataset of 31 small molecules with characterized RNA aptamers[16,26,46,62-88]. Characterized dissociation constants between RNA aptamer-small molecule pairs were converted to binding energies with a blanket assumption of T = 298 K; for each small molecule, the RNA aptamer with the tightest reported binding affinity (lowest K_D) was used to estimate binding energy.

Previous group contribution methods have addressed the identities of functional groups with varying levels of granularity. While many group contribution methods consider as many as 70 different functional groups- some of these methods distinguish a secondary carbon and a secondary carbon in a ring as two different variables[61]- other approaches establish binned functional groups- such as considering all sp^3 carbons as a single variable[56]. While high granularity approaches distinguish among the varying effects of related but different functional groups (primary amines, for instance, might have a contribution quite different from that of secondary amines to many molecular properties), the paucity of small molecule RNA aptamers limits us to using a smaller number of features to avoid overfitting. To this end, we adopted a functional group binning strategy similar to that presented in Andrews et al that resulted in nine separate classes of functional groups[56]. Molecular weight, rotatable bonds, and the presence of aromatic rings were introduced as additional features: prior work has correlated molecular weight with binding affinity and suggested a negative relationship between rotatable bonds and binding affinity, while biochemical analysis of aptamer-target crystal structures has suggested a major role for stacking interactions in aptamer-target binding[16,89].

Not considering every feature meant that we would invariably ignore actual contributions from some functional groups. To work around this, we introduced molecular weight as a variable into our model. We assumed that each functional group would contribute to the overall binding energy of a small molecule in several ways and that beyond polarity and charge, the mass of each group was important as something to bind to. The molecular weight term that we introduced would account for the mass contribution of all the groups in each molecule, thus requiring us only to account for binding energy contributions from polarity and charge.

3.2.3 Multiple linear regression and feature selection

Feature selection was utilized to avoid overfitting and to discard variables that were not predictive for the existing set. Model selection was conducted through an exhaustive all-possible multiple linear regression approach. Two metrics were used to score models: the validation root mean squared error (RMSE) obtained from cross validation and the Bayesian Information Criterion (BIC)[90]. Models were constructed using Python.

A repeated random sub-sampling cross validation was performed with 100 randomly drawn training and test set pairs to estimate the RMSE for each combination of features; in each run, 70% of the data was used for training and 30% for testing. From the 100 subsampled runs, a jackknifed prediction spread for each model was calculated as the average root mean square deviation. 4095 feature combinations covering all possible combinations of the 12 features tested were then sorted by RMSE and BIC (Fig. 3.2a). Each of the 12 features was assessed based on the distribution of ranks of models containing the given variable to determine which variables were more strongly associated with low prediction variance (Supplementary Fig. S3.1).

3.3 Results and Discussion

3.3.1 Multiple regression model for predicting small molecule-RNA aptamer binding energy

This approach yielded a five-feature model for the binding energy ΔG_{bind} of the small molecule with the form:

$$\Delta G_{bind} = 2.71A + 0.011MW + 0.59n_N - 2.49n_{PO_4^{3-}} + 0.33n_{RB} \quad \text{Eq. 3.3}$$

where A indicates the aromaticity of a small molecule, MW is the molecular weight, n_N is the nitrogen count, n_{PO_4} is the phosphate count, n_{RB} is the count of rotatable bonds. As the dataset

used was very small, no separate validation set was set aside at the time of model construction; we accounted for this through repeated random sampling cross validation and an unbiased scoring function. We applied the model on the existing dataset of RNA aptamers and found that nearly three-quarters of the aptamers had characterized binding affinities within an order of magnitude of their predicted K_D (Fig. 3.2b). Through bootstrapping, we found that the standard deviation of the prediction set was 1.15 kcal/mol.

These features were not only present in the one best performing model but were also highly represented in the most predictive models. Three features- molecular weight, aromaticity, and nitrogen count- were represented in at least 70 percent of the top decile of predictive models but in less than 10 percent of the last decile predictive models (Supplementary Fig. S3.1).

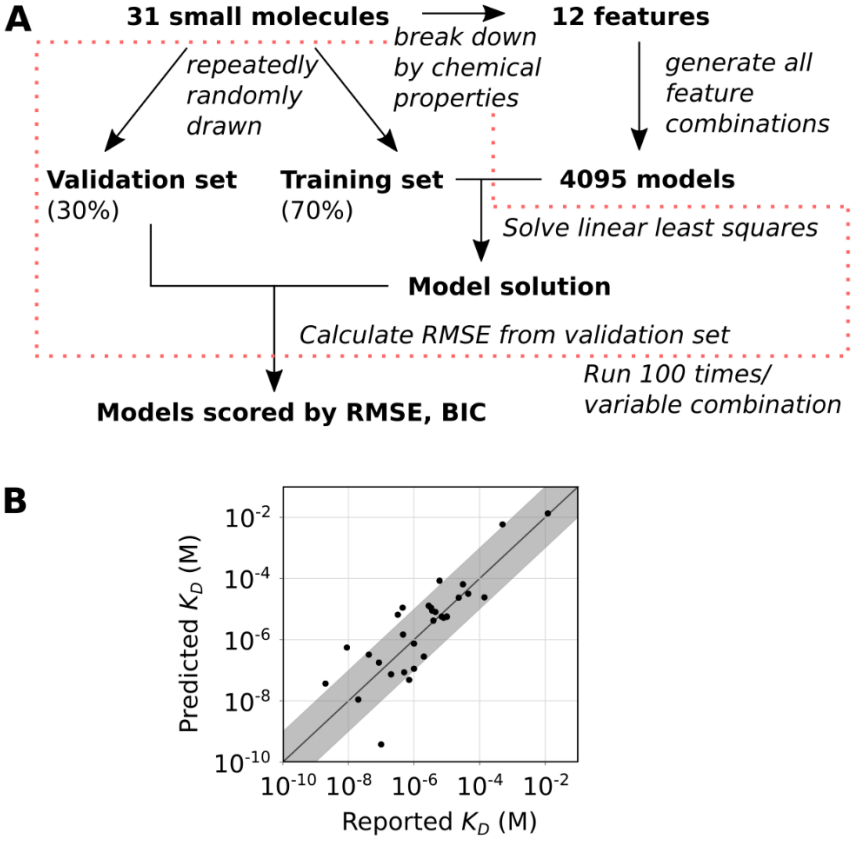


Figure 3.2. Multiple linear regression modeling of small molecule RNA aptamer binding energies. A. In our modeling and cross-validation process, 31 small molecules were described

using 12 features, which were used to generate 4095 models composed of all feature combinations. B. Estimated binding affinities from the model vs. binding affinities reported in the literature. The black diagonal line signifies a one-to-one match, while the dark gray band indicates estimates that fall within one order of magnitude of the reported K_D.

Due to the paucity of data, no test set was reserved from the existing aptamers. However, recent work and outside publications provide three test cases for an initial evaluation of model performance. The *p*-ACA aptamer reported earlier in this work was found to have a K_D of 29 μM, which matches well with the 40 μM predicted K_D. A synthetic naringenin riboswitch was recently reported with an *in vitro* selected naringenin RNA aptamer; although the K_D of the aptamer was not reported, the EC_{50} of the device *in vivo* in *E. coli* was reported between 100-200 μM[91]. As the EC_{50} of aptamer-based devices is expected to be at a higher concentration than the K_D of the aptamer, the measured EC_{50} of the naringenin devices is consistent with the model prediction of a 37 μM RNA aptamer from an *in vitro* selection against naringenin. An RNA aptamer with a K_D of 220 nM was reported for benzylguanine; the model predicted a K_D of 260 nM[92].

The limitations of our approach are similar to the general limitations of group contribution methods: this model is not suited to generating predictions for molecules containing functional groups or chemical characteristics not considered here and there is no prior expectation that it would do well in predicting RNA aptamer binding affinity for classes of molecules that are not represented in the set.

There are many sources of experimental uncertainty in the training data, uncertainty that will still manifest in the characterized binding affinities of small molecule RNA aptamers and thus carry over to our model. I will highlight two major sources of experimental uncertainty and

discuss how to deal with this going forward. The first is uncertainty that arises from variations in aptamer selection approaches. The *in vitro* selections that yielded the aptamers used in the dataset were run under remarkedly varied selection and buffer conditions, using starting pools that varied between 10^{12} and 5 x 10^{15} in size. Utilized buffer conditions exhibited variable pH from 6.5 to 8.5 and buffering agents varied between HEPES, Tris, and PBS, with selections run at temperatures ranging from 4 ºC to 37 ºC[32]. In variable selection conditions, it is likely that RNA folding and the accessible information complexity of RNA sequences will vary, affecting the resultant aptamers. Secondly, binding affinity characterization approaches for selected RNA aptamers also vary. While kinetic approaches such as SPR and BLI have been utilized to yield association and dissociation rates, most quantitative binding assays are thermodynamic, such as equilibrium dialysis, ultrafiltration, and isothermal titration calorimetry[93]. An analysis of different characterization techniques suggests that the characterized K_D can vary by as much as an order of magnitude based on the method used[93]. There are currently no standardized workflows or approaches for this characterization; the predictiveness of this model would likely be improved by such standards or by efforts to characterize existing small molecule aptamers using a consistent technique at consistent characterization conditions.

3.3.2 Biophysical explanation of model parameters

Molecular weight, aromaticity, and the tallies of nitrogens, phosphates, and rotatable bonds were all found to influence small molecule binding energy to RNA. These results largely matched the existing biochemical understanding of how small molecules interact with RNA aptamers. Aromaticity was found to be a large contributor (2.7 kcal/mol) to binding energy, which is consistent with the large role that pi stacking has been reported to play in the stability of

RNA tertiary structure and in the observed interactions of molecules such as theophylline with their cognate aptamers[89]. The calculated thermodynamic contribution of possessing an aromatic ring is consistent with reported thermodynamic contributions of base stacking to RNA stability[94]. Nitrogens in amine groups contribute to binding energy, which is unsurprising given the propensity for hydrogen bonding by amines. Phosphates, which carry a negative charge in the range of physiological pH, have an inverse relation to small molecule binding energy, which is consistent with the expectation that RNA, which itself is negatively charged, will less favorably bind negatively charged targets. The positive relationship between molecular weight and binding energy has been noted in the literature and is further confirmed by this work[16]; this relationship is intuitive, as a larger molecule provides more opportunities for electrostatic interactions.

The positive contribution of rotatable bonds is more difficult to explain. It is a counterintuitive result: more rotatable bonds mean more conformational degrees of freedom and greater entropy, which should make it more energetically unfavorable for a molecule to assume the conformation necessary for interaction with an aptamer. We speculate that this positive rotatable bond contribution instead results from an allowance for internal strain energy. In small molecule-protein interactions, the contribution of internal strain energy to free energy of binding cannot be assessed[95]. The strain energy that a molecule can take is correlated to number of rotatable bonds[96]. Recent work looking at small molecule interactions with proteins suggests that better flexibility for binding can lead to more energetically favorable binding conformations that balance or exceed the strain energy payment and that small changes in conformations can result in large contributions to binding energy[95,97]. Additionally, strain energy in ligand may help pay conformational costs in the substrate- rigid molecules might require too much strain energy from RNA itself and thus penalize binding. As RNA has only four bases and can assume limited

geometries relative to proteins, it is possible that ligand flexibility may provide an even greater benefit here for helping ligands adapt to the fits necessary to maximize bond formation with cognate aptamers.

3.3.3 Analysis of *in vitro* selection parameters

We applied our model for analysis of existing RNA aptamers for small molecules to determine if we could better understand the factors necessary for successful aptamer selections. First, we looked at whether the length of the randomized region of a starting pool was correlated with the residuals between the predicted and reported binding energies of the aptamers in our set (ΔG_{bind} difference); for randomized regions varying from 25 to 120 nucleotides, we saw no trend (Fig. 3.3a). Then, we looked at whether the diversity of the initial selection pool was correlated with ΔG_{bind} difference; as larger pools are intuitively more likely to contain rare, high-affinity solutions, we expected to observe a positive correlation between pool size and aptamers outperforming the model (Fig. 3.3b). Surprisingly, there was no observable trend between these two variables even though initial pool sizes varied between 10^{12} and 5×10^{15}.

We then compared the performance of structured to unstructured RNA aptamer pools: seeding structure such as a stem-tetraloop within the randomized pool is a strategy expected to improve binding affinities. Only three of the selections considered in the set utilized this strategy and all three exhibited tighter than predicted binding affinities.

3.3.4 Potential for applications towards DNA aptamers

We analyzed the performance of the model on predicting binding affinities for small molecule DNA aptamers. The usefulness of applying the modeling approach presented here to

DNA aptamers is apparent, but the number of small molecule DNA aptamers is even smaller than that of small molecule RNA aptamers, making it infeasible to apply the same approach at this time. The differences in binding affinities between RNA and DNA aptamers for the same small molecules suggest that contributions to binding energy from elements of molecular structure may carry different weights dependent on substrate[98-100]. To test the applicability of the model for DNA aptamers, we generated predictions for 12 small molecules meeting the criteria for use with the model that have existing DNA aptamers[98-103]. On average, the RNA model tended to underestimate binding affinity for DNA aptamers: a bootstrapped mean found that small molecule DNA aptamers exhibited a binding energy that was 0.7 kcal/mol higher than the predictions from the RNA model (Fig. 3.3c).

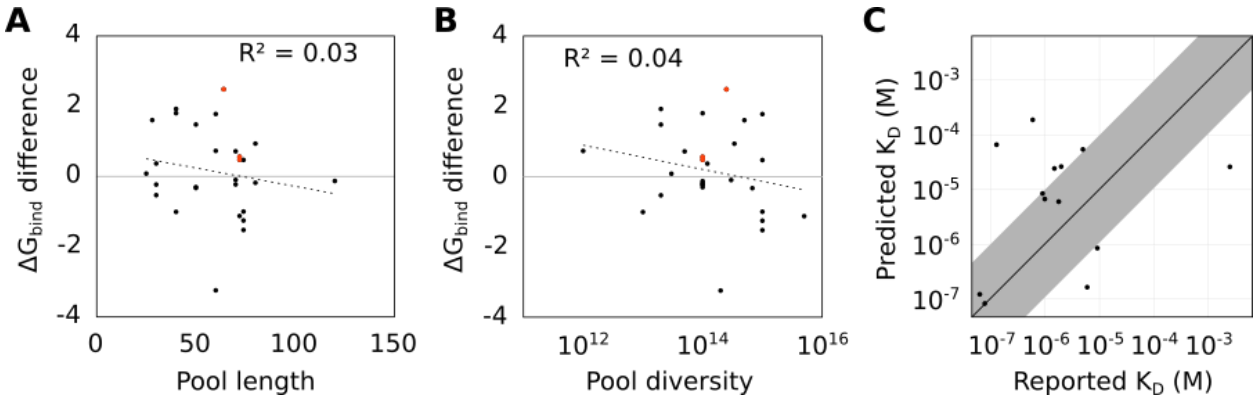


Figure 3.3 Analysis of *in vitro* selections with small molecule binding energy model. A. No trend was found between the length of the randomized region in RNA pools for *in vitro* selections and the performance of resultant aptamers relative to model predictions. Aptamers selected from starting pools with a seeded stem tetraloop are highlighted in red. ΔG_{bind} difference is defined as the measured binding energy of the aptamer minus the predicted binding energy from the model. B. No trend was found between the diversity of the starting pool for *in vitro* selections and the performance of resultant aptamers relative to model predictions. C. 10 of 13

DNA aptamers for small molecules exhibited higher binding energy with their targets than model predictions for RNA aptamers with the same molecules.

3.3.5 RNA Aptamer Metabolite Utility Factor

We have previously discussed the inadequacy of having an aptamer in hand towards building a useful aptamer-based biosensor for a small molecule. The functional concentration ranges for a given aptamer will depend on the intended applications of the aptamer, as different device architectures will have varying effects on the concentrations at which a device built from a particular aptamer can function. These functional concentration ranges in turn are only useful if they overlap with titers of a small molecule that need to be sensed. We will discuss both these criteria more thoroughly and discuss how to quantify the potential utility of an RNA aptamer based off *a priori* predictions using the binding energy model developed in this chapter.

An aptamer is generally not sensitive to concentrations much lower than its K_D. The concentration which an aptamer-based device can sense in turn depends on the architecture used. Theoretically, an aptamer can be useful at concentrations down to its dissociation constant as it can be used as an immobilized biosensor with SPR or BLI or in other direct manners. In practice, this is almost never useful, as such an application is typically expensive and low throughput. Instead, aptamers are typically engineered into a variety of different device architectures; the EC_{50} of devices cannot be more sensitive than the K_D of an aptamer and are typically between two to three orders of magnitude greater than the K_D of the aptamer involved. For instance, a theophylline aptazyme presented by Win and Smolke has an EC_{50} around 1 mM for an aptamer with a characterized K_D around 400 nM, while the pAF4.5 ribosensor has an EC_{50} of 300 µM for an aptamer with a K_D of 3.6 µM[16,18,79,104].

Any application of a sensor will generally have a desired sensitivity- concentrations that the sensor will be expected to distinguish, whether in a quantitative or qualitative manner. We can obtain a semi-quantitative prediction of the potential utility of RNA aptamers for a given metabolic engineering sensing problem by comparing the desired sensitivity for a given application with either the characterized or predicted K_D of an aptamer for that metabolite target. Here, we define the RNA Aptamer Metabolite Utility Factor (MUF), a dimensionless value, as follows:

$$MUF = \log_{10}(S) - \log_{10}(K_D^*) \qquad \text{Eq. 3.4}$$

Where S is the concentration of the metabolite that needs to be detected for an intended application and K_D* is the model generated K_D prediction for a given small molecule. The MUF provides a measure of the distance between the sensitivity necessary for an intended application and the predicted binding affinity of an RNA aptamer for that metabolite, with that quantity reflective of the difference in orders of magnitude between the two values. A positive value for MUF reflects a predicted K_D that is tighter than the needed sensitivity, while a negative value indicates that the predicted K_D insufficient for the application.

As noted, simply having an aptamer K_D at the same concentration range as the sensitivity needed for the intended application is insufficient when building RNA biosensor devices; thus, a small positive value for MUF indicates a low likelihood for the utility of applying an RNA aptamer device in the specified context. MUF values around two or three indicate a higher likelihood of an aptamer being useful in biosensing applications for a metabolite. As discussed, there are many sources of uncertainty in the K_D predictions generated from the model and therefore there are many sources of uncertainty in MUF values; we would not advise using MUF values to inform aptamer selection decisions in the same situations where we advise against

using the model for K_D predictions. As the standard deviation of predictions for the model is 1.15 kcal/mol, all MUF values generated using predictions from the current model will have a spread of ± 0.85.

The desired functional ranges for a sensor can vary for a given molecule at different points in a pathway engineering process, but a reasonable assumption for a target functional sensitivity for a sensor would be the reported maximum titers of a given molecule. While the MUF is ultimately a value that varies based on the specific intended application, we will present generalized MUF values here that reflect the difference between the highest reported biosynthetic titers for a metabolite and the predicted K_D for an RNA aptamer against that metabolite.

3.3.6 Predicting RNA aptamer binding affinities for metabolites in relevant pathways

The usefulness of our model lies in its ability to inform our decisions on what engineered metabolic pathways are amenable to RNA aptamer based biosensing and which metabolites within those pathways are the most promising targets for building biosensors or enacting dynamic genetic control. To determine what sorts of pathways RNA aptamers might best be applied to, we applied our model to generating predictions for industrial chemicals that are currently subject to bio-based production from a comprehensive metabolic map[105]. We generated K_D predictions for some 53 molecules derived from 8 different regions of the metabolic map, comparing chemical product to current bio-based titers to obtain the MUF as well (Fig. 3.4a). From core metabolism, we found that chemicals derived from D-erythrose-4-phosphate were predicted to generate the most high affinity RNA binders and were potentially the most amenable to detection and to implement dynamic genetic control with RNA aptamer-based biosensors.

Individual pathways could be scrutinized as well to pick optimal intermediate metabolites to target for sensing. We chose to analyze four biosynthetic pathways that are currently subject to industrial or academic metabolic engineering research: *p*-AS, naringenin (a flavonoid), resveratrol (a phenylpropanoid), and morphine (an alkaloid)[14,106-108]. In analyzing each pathway, we can see which intermediates are predicted to yield higher affinity aptamers and which are not; by combining this information with an understanding of existing yields of metabolic intermediates, we can make informed decisions when picking small molecules to target with RNA aptamer based biosensing (Fig. 3.4b).

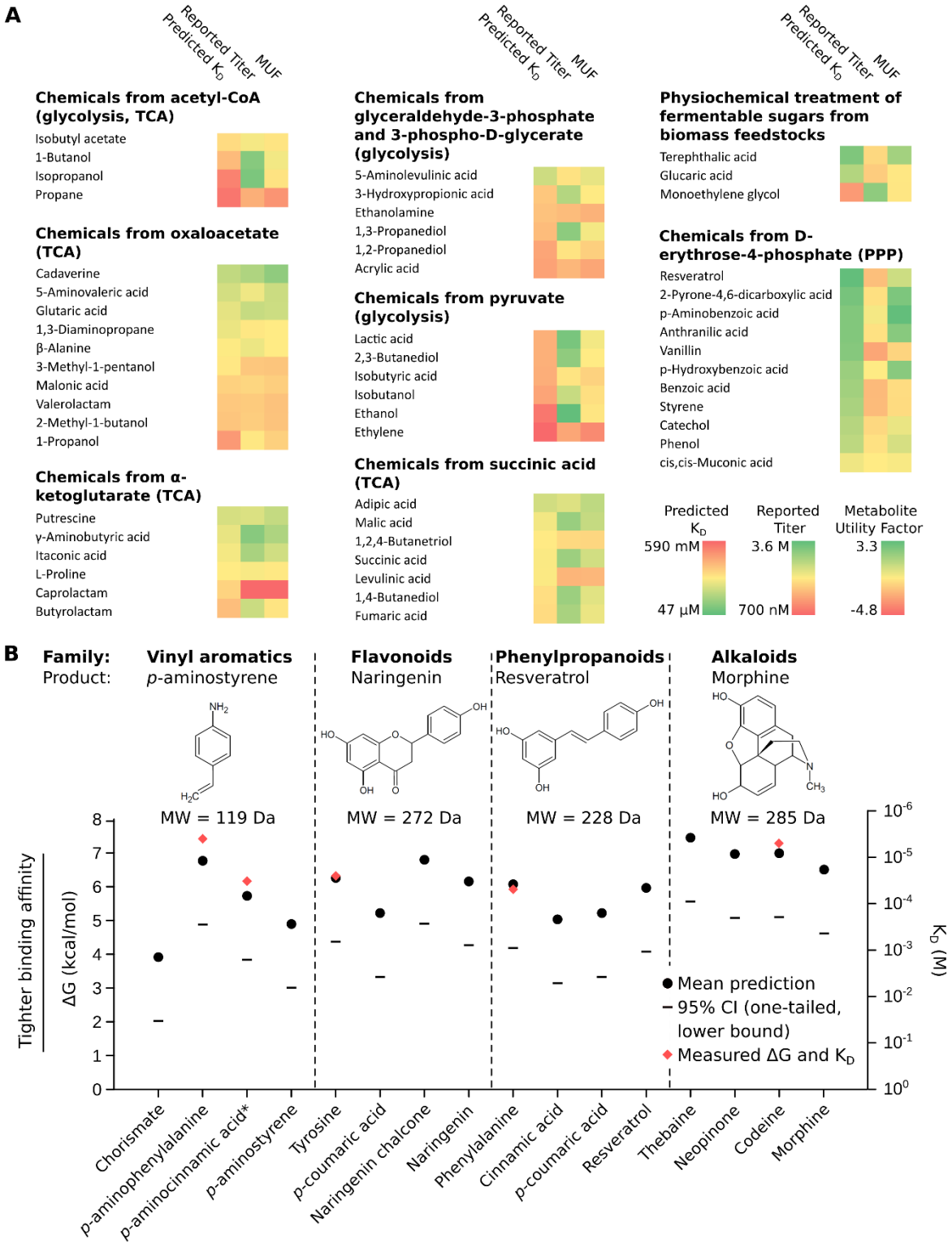


Figure 3.4. Using small molecule binding energy model predictions to pick metabolic pathways and intermediates to target. A. The utility of using RNA aptamers for biosensing in pathway

engineering research and development can be determined by comparing the model predictions for RNA aptamer binding affinities to the final product with the reported titers of the metabolic product of interest. RNA aptamer binding affinities for predicted for fifty-three molecules of interest to metabolic engineering and compared to their reported titers; the difference between the base 10 log of the molar concentration of those values is reported as the metabolite utility factor (MUF), where more positive values of the MUF indicate a higher likelihood of being able to use RNA aptamer biosensors to detect the given final product. B. Model predictions for small molecule targets that are currently subject of pathway engineering work and the intermediate metabolites of those pathways. Characterized binding affinities are shown as well for intermediate metabolites with existing RNA aptamers. Model predictions can inform the appropriateness of using RNA aptamers as biosensors in a given pathway or identify metabolites within pathways that are promising targets for RNA aptamer-based control or sensing.

3.3.7 Conclusions and future directions

This thesis presents the first work at predicting binding affinities for RNA aptamers *a priori* to *in vitro* selections.

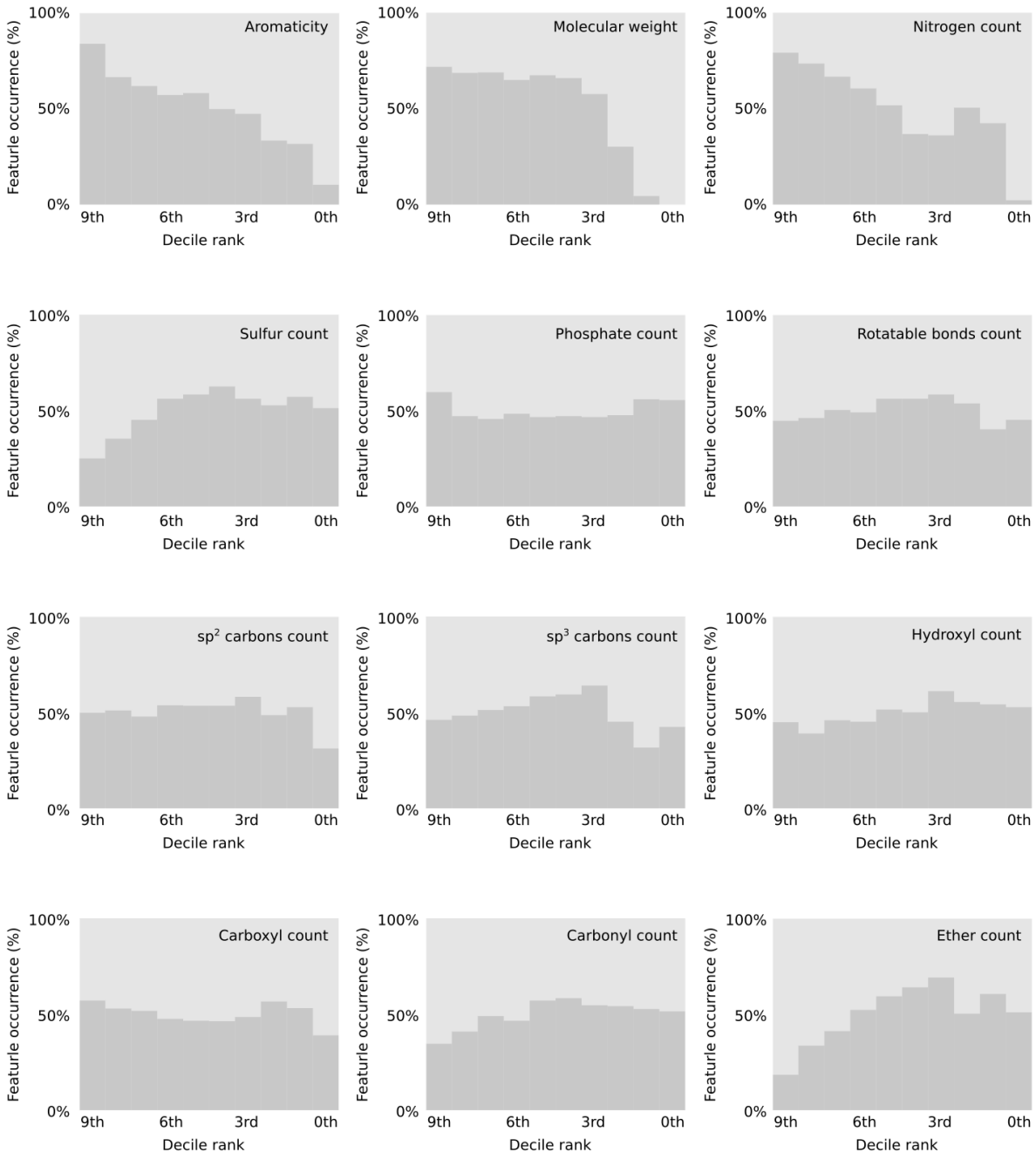
We have demonstrated here an approach for estimating binding affinities of RNA aptamers for small molecules *a priori* based on the properties of the small molecule. However, many applications of RNA aptamers depend not just on the dissociation constant of the aptamer, but on the binding kinetics- the kinetic aptamer ribosensors discussed in this work are one such application. Ultimately, this work would benefit from providing predictions for aptamer binding kinetics- especially association kinetics- in addition to thermodynamic dissociation constants.

The principal challenge to this undertaking is the lack of a large set of characterized small molecule aptamer kinetics. Most studies reporting newly selected aptamers merely report the K_D, characterized through equilibrium approaches such as in-line probing or equilibrium filtration[16,109]. A small number of small molecule RNA aptamers have had their kinetics

characterized by SPR[110]; otherwise there has been no systematic work for characterizing aptamer binding kinetics. To apply a similar modeling approach to predict small molecule aptamer association kinetics, there must be a systematic survey to characterize RNA aptamer binding kinetics; this would likely be best served by using SPR, which is the state of the art for characterizing RNA aptamer kinetics. Once this data is collected, a similar approach to the one detailed here can be used to calculate a regression model to predict aptamer association rates from small molecule chemical properties.

There are also many other applications of this generalized K_D prediction modeling approach. Artificial nucleic acid aptamers are an emerging field that takes advantage of non-standard nucleobases or backbones to access different binding chemistries and conformations, which may allow for tighter binding or greater stability[111-113]. Binding affinity modeling can help us understand whether certain artificial nucleic acid approaches can lead to tighter binding aptamers, and if so, what that binding energy contribution looks like and whether there are certain functional groups or properties that respond to given nucleic acid modifications.

2.4 Supplementary Information



Supplementary Figure S3.1. Distribution of models that incorporate each feature tested in model selection, sorted in ascending order by the mean square error of each model. More predictive models are sorted into higher decile ranks, while the least predictive models have low decile ranks. Variables that are strongly predictive, such as nitrogen count, molecular weight, and aromaticity, are overrepresented in the most predictive models, while variables such as sulfur count are underrepresented in the least predictive models.

Supplementary Table S3.1. Small molecule RNA aptamer dataset used to train model. For each molecule, we used the RNA aptamer with the tightest reported binding affinity and tallied nine different functional groups and three molecular characteristics: aromaticity, molecular weight, and the rotatable bonds count.

Molecule	K_D (M)	Rot Bonds	MW	Aromatic	sp^2	sp^3	N	OH	COOH	CO	PO_4^{2-}	S	-O-
4-aminophenylalanine	3.6E-06	3	180	1	6	2	2	0	1	0	0	0	0
Tyrosine	2.3E-05	3	181	1	6	2	1	1	1	0	0	0	0
Isoleucine	5.0E-04	3	131	0	0	5	1	0	1	0	0	0	0
Phenylalanine	4.5E-05	3	165	1	6	2	1	0	1	0	0	0	0
Theophylline	3.2E-07	0	180	1	3	2	4	0	0	2	0	0	0
Tetramethylrhodamine	8.5E-08	3	389	1	19	4	2	0	1	0	0	0	1
guanosine monophosphate	9.0E-09	4	363	1	4	5	5	2	0	1	1	0	1
adenosine monophosphate	1.0E-06	4	347	1	5	5	5	2	0	0	1	0	1
arginine	1.4E-04	6	174	0	1	4	4	0	1	0	0	0	0
codeine	4.5E-06	1	299	1	8	10	1	1	0	0	0	0	1
flavin mononucleotide	7.1E-07	7	456	1	8	7	4	3	0	2	1	0	2
xanthine	3.3E-06	0	152	1	3	0	4	0	0	2	0	0	0
L-Histidine	8.0E-06	3	155	1	3	2	3	0	1	0	0	0	0
Dopamine	2.8E-06	5	153	1	6	2	1	2	0	0	0	0	0
L-tryptophan	7.0E-06	3	204	1	8	2	2	0	1	0	0	0	0
Valine	1.2E-02	2	117	0	0	4	1	0	1	0	0	0	0
Tobramycin	2.0E-09	6	468	0	0	18	5	5	0	0	0	0	4
Kanamycin	2.0E-07	6	484	0	0	18	4	7	0	0	0	0	4
adenine	1.0E-05	0	135	1	5	0	5	0	0	0	0	0	0
Tetracycline	1.0E-06	2	444	1	10	9	2	5	0	3	0	0	0
biotin	6.0E-06	5	244	0	0	8	2	0	1	1	0	1	0
glutathione	4.2E-08	11	307	0	0	6	3	0	2	4	0	1	0
malachite green	2.0E-06	3	365	1	19	4	2	0	0	0	0	0	0
7-methylguanosine	5.0E-07	2	298	1	4	6	5	3	0	1	0	0	1
s-adenosylhomocysteine	1.0E-07	7	384	1	5	8	6	2	1	0	0	1	1
5-hydroxy-L-tryptophan	3.9E-06	3	220	1	8	2	2	1	1	0	0	0	0
cyclic adenosine monophosphate	1.0E-05	1	329	1	5	5	5	1	0	0	1	0	1
DMHBI	4.6E-07	3	276	1	9	4	2	1	0	1	0	0	1
citrulline	3.1E-05	6	175	0	0	4	3	0	1	1	0	0	0
sphingosylphosphorylcholine	2.0E-08	21	464	0	2	21	2	1	0	0	1	0	0
4,4'-methylenedianiline	4.5E-07	2	198	1	12	1	2	0	0	0	0	0	0

Supplementary Table S3.2. Metabolite Utility Factors calculated for a panel of molecules produced through biosynthesis. For each molecule, we generate a prediction for RNA aptamer binding affinity (K_D*) using our model and compare that affinity with the highest reported biosynthetic titers to calculate the MUF[105].

Molecule Name	K_D* (M)	Titer (M)	Metabolite Utility Factor
Resveratrol	4.66E-05	3.56E-03	1.88
2-Pyrone-4,6-dicarboxylic acid	1.06E-04	9.08E-02	2.93

Terephthalic acid	1.49E-04	4.03E-02	2.43
Anthranilic acid	1.66E-04	1.02E-01	2.79
p-Aminobenzoic acid	1.66E-04	3.14E-01	3.28
Vanillin	1.93E-04	2.96E-04	0.19
p-Hydroxybenzoic acid	4.41E-04	2.65E-01	2.78
Benzoic acid	5.95E-04	2.10E-03	0.55
Styrene	8.32E-04	2.50E-03	0.48
Glucaric acid	1.17E-03	1.23E-02	1.02
Catechol	1.31E-03	4.06E-02	1.49
Phenol	1.77E-03	1.01E-01	1.76
Cadaverine	2.11E-03	1.02E+00	2.68
5-Aminolevulinic acid	3.33E-03	1.12E-01	1.53
Adipic acid	3.88E-03	4.65E-01	2.08
5-Aminovaleric acid	4.32E-03	3.41E-01	1.90
Putrescine	4.82E-03	4.80E-01	2.00
Glutaric acid	8.86E-03	6.81E-01	1.89
γ-Aminobutyric acid	9.88E-03	1.93E+00	2.29
1,3-Diaminopropane	1.10E-02	1.76E-01	1.20
cis,cis-Muconic acid	1.29E-02	2.59E-01	1.30
Malic acid	1.50E-02	1.46E+00	1.99
Itaconic acid	1.62E-02	1.12E+00	1.84
Succinic acid	2.03E-02	1.36E+00	1.83
Levulinic acid	2.10E-02	1.37E-03	-1.19
β-Alanine	2.26E-02	3.63E-01	1.21
L-Proline	2.44E-02	2.94E-01	1.08
1,2,4-Butanetriol	2.53E-02	4.81E-02	0.28
3-Methyl-1-pentanol	2.73E-02	7.77E-03	-0.55
1,4-Butanediol	3.41E-02	1.39E+00	1.61
Isobutyl acetate	3.69E-02	3.10E-01	0.92
Fumaric acid	3.70E-02	1.12E+00	1.48
Caprolactam	4.46E-02	7.03E-07	-4.80
Malonic acid	4.63E-02	3.46E-02	-0.13
Valerolactam	5.79E-02	1.19E-02	-0.69
3-Hydroxypropionic acid	6.01E-02	9.30E-01	1.19
2-Methyl-1-butanol	6.23E-02	1.42E-02	-0.64
Ethanolamine	6.70E-02	2.60E-03	-1.41
Butyrolactam	7.52E-02	6.36E-01	0.93
1,3-Propanediol	7.80E-02	1.77E+00	1.36
1-Butanol	8.10E-02	1.98E+00	1.39

2,3-Butanediol	1.06E-01	1.71E+00	1.21
Lactic acid	1.06E-01	2.50E+00	1.37
Isobutyric acid	1.10E-01	1.33E-01	0.08
1,2-Propanediol	1.37E-01	1.18E-01	-0.06
Isobutanol	1.42E-01	6.75E-01	0.68
Acrylic acid	1.48E-01	1.67E-03	-1.95
Monoethylene glycol	1.78E-01	1.74E+00	0.99
1-Propanol	1.85E-01	1.80E-01	-0.01
Isopropanol	3.26E-01	2.38E+00	0.86
Ethanol	4.23E-01	3.60E+00	0.93
Propane	4.39E-01	7.26E-04	-2.78
Ethylene	5.92E-01	3.17E-04	-3.27

Chapter 4. Probing properties of kinetic RNA aptamer ribosensors for biosensing in metabolic engineering

Abstract

Titratable metabolite-responsive aptamer ribosensors have been demonstrated using a multi-state design strategy; however, underlying biochemical mechanisms for these ribosensors have not been fully elucidated. We demonstrate the utility of the kinetic aptamer ribosensor platform for quantifying yields of the metabolite *p*-aminophenylalanine from supernatants of cultured engineered cells. Through melt and anneal analysis, we demonstrated that we can use RNA folding tools and free energy barrier heights to implement the multi-state designs that exhibit these kinetic behaviors. We demonstrate for the first time a design strategy that utilizes the free energy barriers between RNA structure states to implement kinetic folding traps that greatly increase the achievable sensitivity with a device constructed from a given aptamer. These collective advances showcase the utility of the platform for biosensing in metabolic engineering and a collection of tools for encoding time in kinetic, co-transcriptional RNA devices that can be used to tune sensitivities of aptamer ribosensors.

4.1 Introduction

Kinetic aptamer ribosensors have been reported that utilize the folding paths of *in vitro* transcribed RNA aptamer-based devices to create metabolite-responsive fluorescence outputs[18]. This architecture has been demonstrated to yield *p*-AF responsive devices that bind kinetically-that is, in a regime controlled by the kinetics of association and dissociation rather than by the binding affinity[18]. Designed ribosensors have already demonstrated ligand-responsive titratable behavior in response to *p*-AF in assays that can be run *in vitro* with many samples in parallel in just an hour.

p-AF devices have exhibited EC_{50} as low as 300 μM and are sensitive to concentrations below 100 μM. In current work on the *p*-AS pathway in the Carothers Group, *p*-AF titers of about 200 mg/L (~ 1 mM) have been achieved, which places culture supernatant in the range easily detectable by the *p*-AF aptamer ribosensors, making this a plausible approach for high throughput screening. However, to be useful tool for high throughput metabolite screening, kinetic aptamer ribosensors must demonstrate an ability to quantify titers from cultured cell supernatants in a manner comparable to existing analytical chemistry techniques. Here, we use the *p*-AF aptamer ribosensors to quantify *p*-AF from engineered cells and demonstrate comparable results to HPLC.

The multi-state design of kinetic aptamer ribosensors is based on an engineered folding trajectory through a set of structure states, dictated by setting the free energy barriers between each set of structure states (Fig. 4.1). Seven structure states were defined for RNA folding analysis: one incomplete transcript with the aptamer transcribed ($s1_{co\text{-}txn}$) and six other structure states encompassing the full sequence of the ribosensor. State $s1_{co\text{-}txn}$ represents a ligand competent state in which the toehold actuator has not yet been transcribed, but the aptamer is

able to bind to and become stabilized by its target ligand. State $s2_{full}$ results from $s1_{co\text{-}txn}$ at the conclusion of individual *in vitro* transcription events and represents the ligand competent state in which the device is available for binding to its target ligand and transitioning to structure state $s3_{full}$; the binding window, or t_{bind}, is available for the time encompassing $s1_{co\text{-}txn}$ and $s2_{full}$, or the full time that the aptamer is in a ligand competent state.

Structure state $s2_{full}$ is separated from state $s3_{full}$ representing the actuation state by free energy barrier B1 and from state $s4_{full}$ representing the non-signaling off state by B2. By design, the magnitude of barrier B1 is greater than that of B2, favoring a transition of $s2_{full}$ into the minimal free energy $s4_{full}$ off state in the absence of ligand. In the presence of ligand, metabolite binding to the aptamer stabilizes the ligand competent state and reduces the energy barrier necessary to transition to the actuation state, favoring that transition instead.

The off state $s4_{full}$ can transition to either a background leak state $s5_{full}$ through energy barrier B3, in which the aptamer does not refold but the toehold is made available for strand displacement, or to a ligand-dependent thermodynamic state $s6_{full}$ through energy barrier B4, in which the aptamer refolds and the ligand competent device can bind the ligand thermodynamically. Both barriers B3 and B4 are maximized to prevent non-specific, non-kinetic signal, expect in the case of thermodynamic devices, in which B4 is minimized (Supplementary Table S4.1).

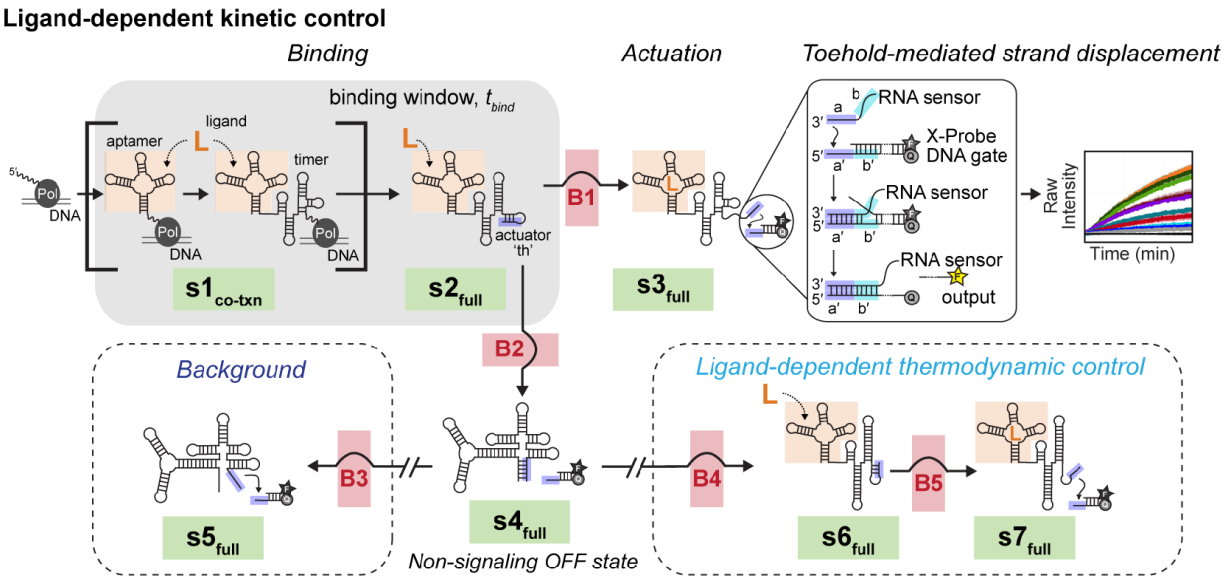


Figure 4.1. Multi-state design for kinetic aptamer ribosensors. Secondary structure states are indicated in green while free energy barriers separating states are indicated in red. Ribosensors are available for aptamer-ligand interaction during the binding window t_{bind}, which is available co-transcriptionally as the structure state $s2_{full}$. Kinetic ribosensor design screens for sequences with correct rank order and ranges for free energy barriers B1, B2, B3, and B4 to provide a reliable RNA folding path through the structure states. The activated state $s3_{full}$ generates a ligand-dependent response by interaction of the toehold actuator with a DNA gate. Adapted with permission from Burke (2017).

Two families of *p*-AF aptamer ribosensors have been designed: the first, represented by the ligand-sensitive and titratable pAF4.5, pAF4.9, and pAF4.11, is designed using kinetic multi-state barrier height analysis, while the second, represented by pAF1.4, is designed to act thermodynamically without free energy barriers B1 and B2 separating $s2_{full}$ from either the actuation $s3_{full}$ state and the off $s4_{full}$ state. As the off state remains the MFE state for the full sequence of pAF1.4, ribosensors preferentially turn off, but combined with a lower B4 barrier height (Supplementary Table S4.1), this device can access the thermodynamic activation $s6_{full}$ state.

To validate that barrier height design can distinguish between thermodynamic and kinetic devices, an assay was developed to characterize the kinetic contributions to aptamer ribosensor signal generation. Once fully transcribed, aptamer ribosensors are expected to fold into the non-actuating off state ($s4_{full}$) in a matter of seconds or less; unless engineered to act thermodynamically, these devices are expected to generate minimal ligand-dependent signal upon reaching the $s4_{full}$. To assess this, *in vitro* ribosensor assays with *p*-AF responsive devices were spiked with Proteinase K early to end transcription; signal dependence on ligand was then compared (Supplementary Fig. S4.1). pAF1.4 was found to exhibit about a 90% thermodynamic character, while the three kinetic *p*-AF devices all exhibited >90% kinetic responses[18].

RNA sequences are capable of assuming many stable conformations short of the minimum free energy state[114]. Metastable folding states, which can accessed when RNA sequences fold into a conformation that is thermodynamically favored by a partial transcript although the full transcript has a different MFE structure and becomes kinetically trapped, can result from high free energy barriers between the trapped state and the MFE state[115].

While kinetic traps can potentially frustrate RNA folding to a designed MFE state, they can also inversely be utilized to slow folding into an MFE state. Thus, metastable folding states can be coopted to temporarily maintain structure states at local energy minima before restructuring to the lowest energy state. We hypothesize that metastable folding states can be engineered by screening for sequences with large free energy barriers between the co-transcriptional trapped state and the MFE state. In the context of kinetic aptamer ribosensors, metastable folding states may offer a straightforward method of extending the binding window t_{bind} when the ligand competent $s2_{full}$ is available before refolding into the off state $s4_{full}$, thus improving the achievable sensitivity of a device using a given aptamer. This folding state would

thus be governed by the lowest free energy barrier height separating this trapped structure from refolding into another structure, which in this case would be the free energy barrier B2 separating $s2_{full}$ from $s4_{full}$. Here, we demonstrate through a binding time analysis that the pAF4.5, pAF4.9, and pAF4.11 ribosensors exhibit metastable folding states to achieve the comparatively high sensitivities exhibited by those devices.

Aptamer ribosensors have previously demonstrated titratable *in vitro* behavior using multi-state folding design with co-transcriptional assays. We demonstrate here that this platform can be used to quantify metabolite titers from engineered *E. coli,* making this a potentially valuable tool for high-throughput screening of engineered cells. We demonstrate that this utility derives from kinetic activity of the sensors and that we are capable of engineering kinetic versus thermodynamic activity in the sensors through our multi-state folding design. We additionally demonstrate that our free energy barrier height analysis for multi-state design has allowed us to engineer kinetic folding traps for these co-transcriptional devices, the first demonstration of a designed use of this common natural phenomenon.

4.2 Materials and Methods

4.2.1 DNA gate preparation

DNA gates were either two-stranded or 4-stranded X-probe gates; for two-stranded gates, 40 μM of each strand was combined with 10 μM $MgCl_2$ and for X-probe gates, 20 μM of each strand was combined with 10 μM $MgCl_2$. Oligos- including fluorophore and quencher modified oligos- were ordered from Integrated DNA Technologies (IDT). DNA gate reactions were annealed by heating to 95^0 C for 1 minute and then dropping by 1^0 C per minute down to 25^0 C;

gates were then purified on a 10% native polyacrylamide gel, which was run for 30-45 minutes at 3 W. Gates were separated from unannealed oligos and cut from the gel after UV shadowing and gel slices were incubated in 300 μL of 300 mM KCl overnight to elute the DNA gates. Eluted gates were then cleaned up via spin filters and an ethanol precipitation with GlycoBlue. The gates were resuspended in 10 μM Tris.

4.2.2 Ribosensor DNA template preparation

Ribosensor template oligos were ordered as ultramers from IDT. Double-stranded DNA templates were assembled with 1 nM of the template and 1 μM each of forward and reverse primers; the forward primer appended the T7 promoter. PCR was performed with DreamTaq 2x Master Mix. The resulting product was then diluted a thousand-fold and underwent PCR again with 1 μM primers; this was cleaned up using a Qiagen PCR clean-up kit.

4.2.3 Ribosensor assays

Ribosensor assays were run on an Agilent Mx3005p qPCR machine using white PCR tubes with clear caps. Each reaction was 20 μL, containing 150 mM potassium glutamate, 5 mM DTT, 1mM each rNTP, NEB T7 transcription buffer, 50 U T7 RNA polymerase, 30 ng ribosensor template, 5 ng control template, and 100 nM each of the ribosensor and control gates. Triplicate reactions were assembled on ice and then measured in the qPCR machine at 37^0 C every 15 seconds with the ROX and the FAM channels.

In the production assays, we centrifuged cultures at 4000 g for 15 minutes and separated out the supernatant; 10 μL of this supernatant went into the 20 μL ribosensor assays, which used 6 mM $MgSO_4$ in place of the NEB T7 transcription buffer used in the *in vitro* assays.

p-aminophenylalanine and tryptophan were both made up fresh for relevant experiments.

4.2.4 Melt and anneal assay

Ribosensor RNA was transcribed overnight from ribosensor DNA template using T7 RNA polymerase at 37^0 C with 5 mM each rNTP, 10 mM DTT, 0.01% Triton-X 100, 25 mM MgCl2, RiboLock, T7 RNA polymerase, 2.5 mM Spermidine, and 40 mM Tris pH 7.5. RNA was gel purified using denaturing urea 8% polyacrylamide gel run at 15 W for 70 min. Ribosensor RNA bands were cut from the gel and RNA was extracted from gel slices by a 2-hour electrophoresis at 200 V using a Whatman Elutrap. RNA was then ethanol precipitated and resuspended in water. RNA melt-and-anneal was run in an annealing buffer consisting of 10 mM Tris pH 7.5, 50 mM NaCl, and 1 mM EDTA, with RNA heated in annealing buffer to 85^0 C for 1 min and then cooled to room temperature (25^0 C) with a 1^0 C drop every 30 seconds using a thermocycler.

Thermodynamic ribosensor reactions were assembled with 2 pmole gate per reaction; for each thermodynamic reaction, two reaction conditions were tested with 2:1 and 8:1 ratios of ribosensor RNA to ribosensor gate. Reactions were run in 150 mM potassium glutamate, NEB T7 reaction buffer, and 5 mM DTT. Reactions were mixed at 37^0 C and then immediately transferred to a qPCR machine; ROX fluorescence was measured at 15 second intervals for 1 hour.

4.2.5 Lowered rNTP concentration ribosensor assays

Lowered rNTP concentration *in vitro* ribosensor assays were set up largely the same was as standard in vitro ribosensor assays, save for a drop to 0.1 mM each rNTP in the reaction (Each

reaction contained 150 mM potassium glutamate, 5 mM DTT, NEB T7 transcription buffer, 50 U T7 RNA polymerase, 30 ng ribosensor template, 5 ng control template, and 100 nM each of the ribosensor and control gates).

4.2.6 Proteinase K ribosensor assays

Proteinase K ribosensor assays required preparation of standard in vitro ribosensor assay 20 μL reactions, with 150 mM potassium glutamate, 5 mM DTT, NEB T7 transcription buffer, 50 U T7 RNA polymerase, 30 ng ribosensor template, 5 ng control template, and 100 nM each of the ribosensor and control gates combined on ice in white, clear-capped PCR tubes. Proteinase K was prepared from dry powder (Invitrogen) at a concentration of 1200 units/mL in a buffer of 50% glycerol and 5 mM Tris pH 7.4. 2.4 units of Proteinase K were added per reaction around 11.25 minutes to halt transcription and *p*-AF was spiked into the no-ligand conditions at 60 minutes to observe thermodynamic signal. Assays were run for 90 minutes total.

4.2.7 *E. coli* RNA polymerase transcription assays

pAF4.9 was transcribed using both T7 RNA polymerase and *E. coli* RNA polymerase to compare transcription performance from each approach. For transcription from *E. coli* RNA polymerase, the pAF4.9 ribosensor was rebuilt and the 5' T7 promoter was replaced with a p70a promoter with 40 additional bases of sequence 5' of the promoter[17]. Each reaction was 50 μL; the T7 RNA polymerase transcription reaction contained 150 mM potassium glutamate, 1 mM each rNTP, NEB T7 transcription buffer, 125 U NEB T7 RNA polymerase, 40 U Thermo Scientific RiboLock RNase Inhibitor, and 75 ng ribosensor template with T7 promoter. The *E. coli* RNA polymerase transcription reaction contained 1 mM each rNTP, NEB *E. coli* RNA polymerase

transcription buffer, 10 U NEB *E. coli* RNA polymerase Holoenzyme, 40 U Thermo Scientific RiboLock RNase Inhibitor, and 250 ng ribosensor template with p70a promoter.

Reactions were assembled at 37 ^{0}C and then initiated with the addition of the appropriate polymerase at time 0 minutes. Time points were collected at 5, 20, 60, and 240 minutes by quenching 10 μL of the reaction with 20 μL of formamide EDTA loading buffer and storing immediately at -80 ^{0}C. 10 μL of each quenched reaction were then heated to 90 ^{0}C for 2 minutes, then set on ice for 5 minutes, then loaded and run out on a 0.75 mm 8% polyacrylamide gel for 1.5 hours at 7 W. The resulting gel was stained with Sybr Gold (Invitrogen) in 0.5X TBE for 10 minutes and imaged using a BioRad PharosFX Molecular Imager. Gel images were analyzed and RNA bands on the gel were quantified using ImageJ; for each transcription condition, the four time points were fit with a first order rate equation using Microsoft Excel.

4.3 Results and Discussion

4.3.1 Analyzing metabolic production with kinetic aptamer ribosensors

While *p*-AF ribosensors have been shown to respond in a titratable manner to *p*-AF in both *in vitro* conditions and in spent cell culture media from spiked *p*-AF, the utility of these sensors lie in their ability to quantify yields from cell supernatants from *p*-AF producing engineered *E. coli*. This was tested with an engineered *E. coli* transformed with plasmids carrying the PapABC operon under Tetracycline-inducible control and enhancing the activity of endogenous *aroG* and *aroL*. Varying levels of *p*-AF production were achieved using varying amounts of anhydrotetracycline (ATc) to induce expression of PapABC (Fig. 4.2a). Comparison of reported titers from cell culture supernatants using HPLC and the pAF4.9 ribosensor showed a

$R^2 = 0.81$ correlation between the two approaches. At low to medium levels of ATc induction (0 nM ATc to 45 nM ATc), the two approaches showed strong correlation ($R^2 = 0.98$), suggesting that when run in triplicate, the pAF4.9 ribosensor was appropriate for quantification of *p*-AF titers (Fig. 4.2b).

At higher levels of PapABC induction, the values between the two approaches diverged, with the ribosensor providing higher titer readings at 60 nM and 75 nM ATc. The cause of this divergence is still not known, though we propose that it is a result of interactions between the *p*-AF aptamer and other metabolites that are responsive to levels of PapABC expression. We probed the possibility that the pAF4.9 ribosensor was responsive to *p*-aminophenylpyruvic acid (*p*-APPA), a metabolic intermediate that results from enzymatic conversion of chorismate by the expressed PapABC operon and is converted to *p*-AF by expression of endogenous *tyrB*. Insufficient flux through *tyrB* could presumably cause a buildup of *p*-APPA at high expression of PapABC, which might explain the mismatched response of the pAF4.9 ribosensors for high induction of PapABC.

To probe this, we tested whether the pAF4.9 ribosensor responds to *p*-APPA. Our *in vitro* assays titrating p-APPA for the pAF4.9 ribosensor showed that the ribosensor was not responsive to *p*-APPA at 1 mM of the ligand; higher concentrations of *p*-APPA created a transcription defect in the assays but did not suggest that the ribosensor was responsive to *p*-APPA (Supplementary Fig. S4.2). This result is not consistent with the hypothesis that pAF4.9 ribosensor signal mismatch with HPLC is generated by increased levels of *p*-APPA at high induction of PapABC. Despite this signal mismatch at higher levels of enzyme activity, we have demonstrated that the pAF4.9 ribosensor can provide semi-quantitative reads on *p*-AF yields and that at the very least it is capable of distinguishing between the presence and absence of *p*-AF.

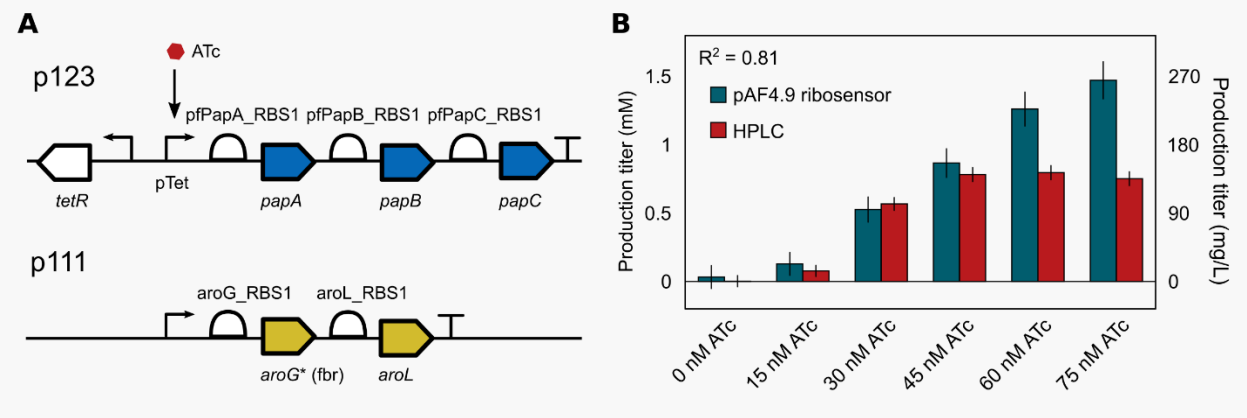


Figure 4.2. Quantification of *p*-AF yields from engineered *E. coli* using the pAF4.9 ribosensor. A. A Strain of *E. coli* was engineered for *p*-AF production by transformation of plasmids heterologously expressing the PapABC operon and enhancing endogenous expression of *aroG* and *aroL*. PapABC expression was inducible by ATc from a Tet promoter. B. Comparison of production titers measured by the pAF4.9 ribosensor and HPLC. At low to medium ATc induction, measured production titers align well between the two methods; overall R^2 between the two methods is 0.81.

4.3.2 Demonstrating titratable response from the aromatic amino acid ribosensor

Initial work on the kinetic aptamer ribosensor platform demonstrated titratable fluorescent responses from devices engineered with the *p*-AF aptamer. Later work demonstrated that the tyrosine aptamer- which has been shown to interact with varying degrees of sensitivity to a panel of aromatic amino acids, including tryptophan, tyrosine, phenylalanine, and L-DOPA- could be incorporated into a RNA aptamer ribosensor responsive to aromatic amino acids[18]. We probed whether this titratable behavior of kinetic ribosensors extended to aptamers other than the *p*-AF aptamer by running titrations of tryptophan and tyrosine, to which the aromatic amino acid ribosensor had been most responsive at 1 mM, with the aromatic amino acid ribosensor constructed from the tyrosine aptamer. Due to the aqueous solubility limits of tyrosine (<3 mM

solubility in water at 25° C), we were not able to observe titratable behavior of the ribosensor with that molecule. However, we were able to demonstrate titratable behavior with an EC_{50} of 4.1 mM using tryptophan, which has a higher aqueous solubility (Fig. 4.3).

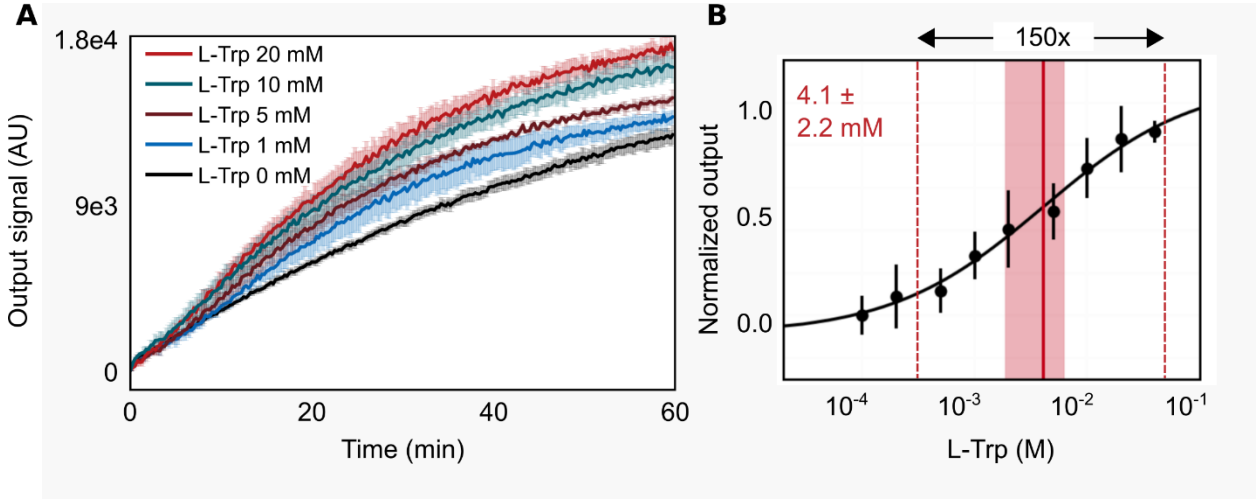


Figure 4.3. Titratable behavior of the aromatic amino acid ribosensor in response to L-tryptophan. A. Kinetic traces from co-transcriptional ribosensor experiments demonstrating increasing fluorescent signal generation from increasing concentrations of tryptophan. B. Concentration sensitivity of the aromatic acid ribosensor responding to tryptophan. Output is normalized to maximum of Hill function fit. Tryptophan was added in concentrations ranging from 100 µM to 40 mM. The EC_{50} of the ribosensor to tryptophan was found to be 4.1±2.2 mM and the usable dynamic range is indicated by the dotted lines.

4.3.3 Demonstrating kinetic vs. thermodynamic activity in aptamer ribosensors

Kinetic traps can potentially confound multistate RNA folding design, as RNA devices can fold into thermodynamically suboptimal structures and become trapped without reaching the specified MFE states that are the basis of aptamer ribosensor design. The Proteinase K assay approach is co-transcriptional and thus does not address the question of whether thermodynamic and kinetic action is occurring as a result of our multistate design. Thus, it would be instructive

to demonstrate that the MFE structure for pAF1.4 maintains activity and thus that pAF1.4 is a thermodynamic device, while showing that the MFE structures for pAF4.5, pAF4.9, and pAF4.11 have minimal activity, showing that these are kinetic devices.

To test whether the results from multistate design seen from the Proteinase K assay were a result of the MFE structure of each device, we ran a melt and anneal assay. In this assay, devices were transcribed, then melted and slowly annealed to drive folding into the MFE state. Fluorescent output was then observed from ribosensor devices in the presence and absence of ligand. We found that pAF1.4, a thermodynamic device, generated a much larger absolute signal in the presence of ligand than the three kinetics devices (Fig. 4.4a); additionally, we found that melt-and-anneal pAF1.4 exhibited a higher activation (tenfold) in the presence of ligand than melt-and-anneal kinetic devices, none of which were activated by more than twofold in the presence of ligand (Fig. 4.4b).

As pAF1.4 is designed with a substantially lower free energy barrier for B4 (the barrier for the transition between the MFE off state and the thermodynamically ligand competent $s6_{full}$ state, the higher ligand activation of pAF1.4 supports our hypothesis that free energy barrier height design can help govern the rate of transitions between states and thus the mode of action (thermodynamic vs. kinetic) of an RNA device. A melt and anneal assay run on the aromatic amino acid ribosensor demonstrated no difference in signal between the ligand present and absent cases, suggesting that the aromatic amino acid also primarily functioned as a kinetic rather than as a thermodynamic device (Supplementary Fig. S4.3).

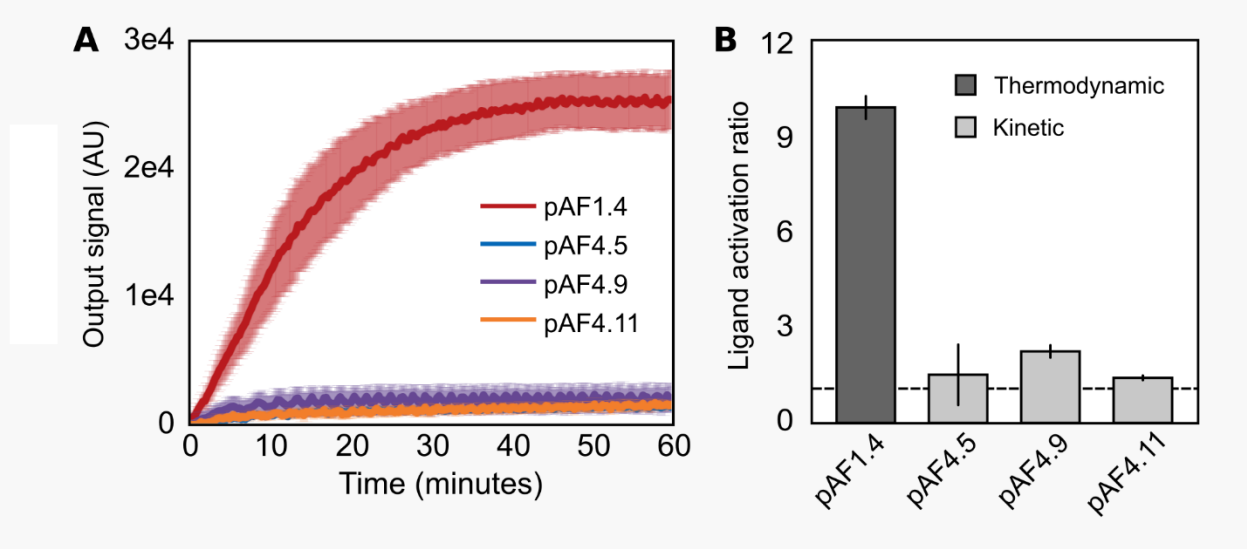


Figure 4.4. Thermodynamic and kinetic devices exhibited differing ligand-dependent responses in a melt and anneal assay. A. pAF1.4, a thermodynamic device, exhibited strong activation by *p*-AF after melt and anneal, while the kinetic devices exhibited minimally elevated signal. B. pAF1.4, a thermodynamic device, exhibited a tenfold activation in the presence of ligand, contrasting strongly with the kinetic devices, none of which exceeded a twofold activation in the presence of *p*-AF in the melt and anneal assay.

4.3.4 Utilizing metastable folding states to encode sensitivity of co-transcriptional RNA devices

So far, we have primarily drawn a distinction between kinetic and thermodynamic binding of ribosensors, governed by free energy barrier B2. However, previously discussed metastable state mechanisms can create two separate modes of kinetic binding: co-transcriptional and post-transcriptional. Co-transcriptional binding results in activation of the ribosensor by ligand while the states $s1_{co\text{-}txn}$ and $s2_{full}$ are available during transcription, which is governed by elongation time. Post-transcriptional binding results in activation of the ribosensor by ligand while $s2_{full}$ is available after transcription and before the ribosensor has rearranged into the MFE off state $s4_{full}$, which is governed by free energy barrier B2. Post-transcriptional binding would

only be able to occur if the transcribed RNA is kinetically trapped in $s2_{full}$ and does not automatically rearrange to $s4_{full}$.

Other discussed forms of sensitivity tuning such as timer length and polymerase elongation rate during transcription are co-transcriptional in nature, making the shifts in sensitivity ultimately dependent on sequence length, reaction conditions, and the choice of polymerase. The limitations of timer length are manifest: as device design is predicated on structural rearrangement and computational folding design, it becomes increasingly difficult to manage both refolding of a longer sequence in three-dimensional space and the computational power necessary to assess the thermodynamics of longer sequences; when dealing with timers that are fewer than 100 bases, modifying timer length simply changes t_{bind} by fractions of a second when T7 RNA polymerase elongation rate is 200 nucleotides per second[117]. On the other hand, kinetic traps occurring in nature are known to resolve on the time scale of seconds or longer[115].

In previous work, we've characterized the kinetics of the *p*-AF aptamer using biolayer interferometry, with findings that align with earlier thermodynamic characterizations of the aptamer. For the *p*-AF aptamer, we found a k_{on} of 490±15 $M^{-1}s^{-1}$ and a k_{off} of 3.9±0.1x10^{-3} s^{-1}, with a K_D of 7.9±0.2 µM, which was consistent with the literature report of a 3.6 µM K_D and the general spread among varied forms of aptamer dissociation constant characterization[1,19].

We estimated the co-transcriptional binding window of each device using literature values for T7 RNA polymerase elongation and the timer length of each device[117] (Supplementary Table S4.2). We used the measured EC_{50}s and the Eq. 1.1 to calculate the actual available binding time. From this, we find that while the measured and calculated binding times align well for pAF1.4, measured and calculated binding times differ greatly for the pAF4 generation

devices, with differences of over 10-fold (and 4 seconds) for pAF4.5 (Fig. 4.5a). We interpret this difference between the t_{bind} calculated from measured EC_{50}s and the predicted co-transcriptional t_{bind} as the binding time encoded in a metastable state; we define the metastable state time as $t_{meta,pred} = t_{bind\ measured} - t_{co\text{-}txn}$.

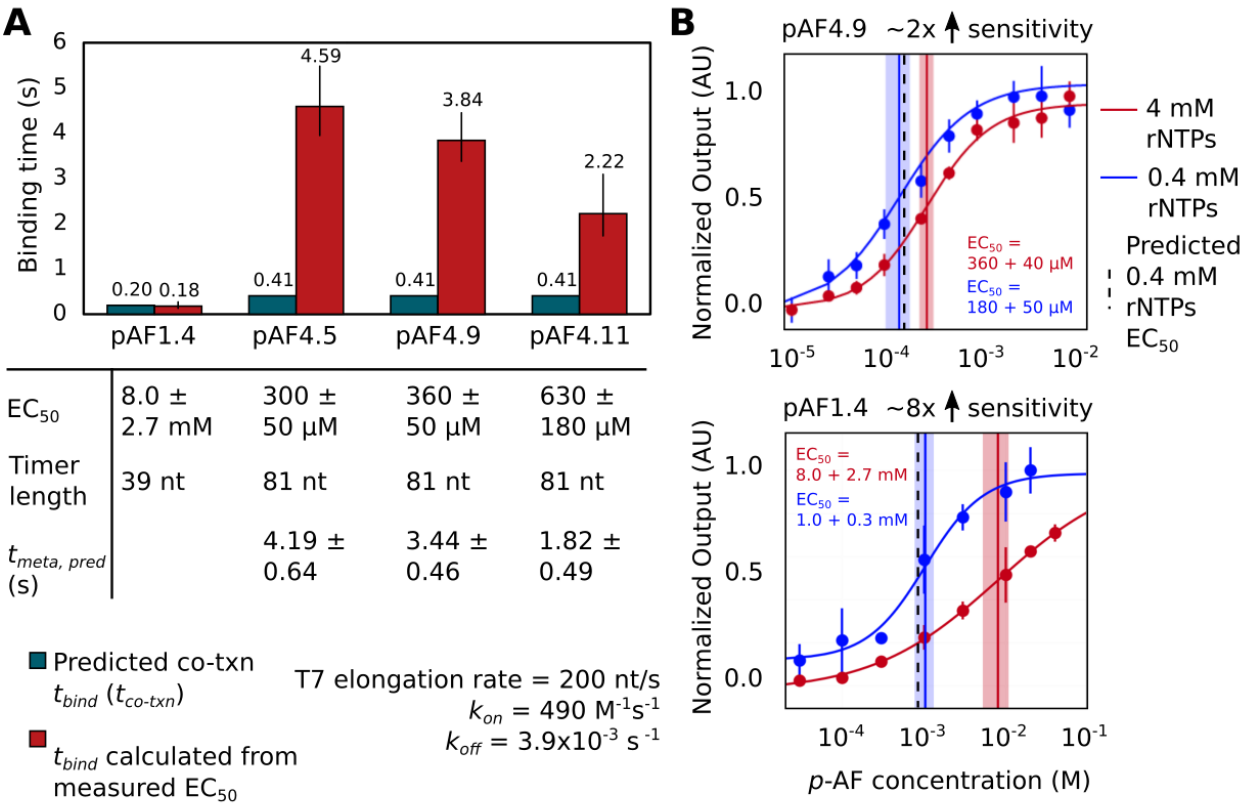


Figure 4.5. Metastable states provide additional binding time for kinetic devices. A. Comparison of predicted co-transcriptional binding times from timer lengths to device binding times calculated from EC_{50}s for pAF1.4 and three pAF4 generation devices. While the binding time from the pAF1.4 EC_{50} corroborates the predicted co-transcriptional binding time, there are large differences for the pAF4 generation devices. B. *In vitro* titrations of pAF4.9 and pAF1.4 under normal and low rNTP conditions; output is normalized to maximum value from Hill function fit. Changing the assay rNTP concentration from 4 mM to 0.4 mM reduces the RNA polymerase elongation rate, which increases the duration of the binding window (t_{bind}) and shifts device EC_{50}s to lower concentrations of *p*-AF. The observed device sensitivity change is higher for pAF1.4, a device engineered without a metastable state, than for pAF4.9, which has free energy

barriers that specify an extended t_{bind} through metastable states. The larger magnitude EC_{50} change for pAF1.4 suggests that pAF1.4 is more sensitive to changes in co-transcriptional binding time, which is consistent with hypothesis that the t_{bind} for pAF4.9 is partially encoded by metastable folding states that are not affected by co-transcriptional binding time.

We expect that the metastable state time $t_{meta,pred}$ would not vary based on transcription rate, as the full-length transcript, once-folded, should refold into the MFE at the same rate regardless of the previous elongation rate; this refolding rate is governed by the barrier height specifying that structure state. If the co-transcriptional binding time is changed by slowing the elongation rate, we should see a larger change in sensitivity for pAF1.4 than for the pAF4 generation devices, as the t_{bind} for those devices is already primarily governed by t_{kt}. Using the calculated metastable state time for each device and assuming that a 10-fold drop in rNTP concentration would result in a 10-fold slowdown of T7 polymerase elongation, we predicted the EC_{50}s for both pAF1.4 and pAF4.9 at low rNTP conditions to be 730 μM and 190 ± 20 μM. pAF1.4 and pAF4.9 were both tested at low rNTP concentrations to slow elongation and test these predictions hypothesis; at 10-fold reduced rNTP concentrations, pAF1.4 and pAF4.9 exhibited EC_{50}s of 1.0 ± 0.3 mM and 180 ± 50 μM, respectively, which aligned well with our predictions when considering metastable state time encoding (Fig. 4.5b).

To our knowledge, this is the first demonstration of designed RNA metastable folding states for time encoding to achieve increased device sensitivity. The advantages of kinetic devices, the importance of binding time windows in kinetic devices, the role of free energy barriers in creating RNA metastable states, and the role that kinetic traps can play in maintaining thermodynamically suboptimal RNA structures for extended time windows are already understood. However, this is the first work to collectively synthesize these elements and

demonstrate that an engineering approach through free energy barrier analysis can implement kinetic folding traps to improve the sensitivity of kinetic RNA devices. Kinetic folding traps, once seen as a bug- a hindrance to correctly folding large, complex RNAs- are here turned into a feature, allowing us to tune the sensitivity achievable by kinetic RNA aptamer ribosensors.

4.3.5 Future directions for tuning ribosensor sensitivities

There remain many unexplored ideas for further tuning the sensitivity of kinetic aptamer ribosensors by extending the binding window t_{bind}. The current platform utilizes T7 RNA polymerase to transcribe the described RNA devices. An alternative approach that would increase elongation time is switching from a T7 RNA polymerase platform to one run using *E. coli* RNA polymerase. *E. coli* RNA polymerase elongation rates have been reported around 20-25 bases per second, substantially slower than the 50-230 base per second elongation rates reported for T7 RNA polymerase[117-119]. This slower elongation rate would provide more time for co-transcriptional ligand activation of the ribosensors and we would expect to see this reflected as an increased sensitivity for these devices (Supplementary Fig. S4.4a). This approach might drop elongation rate similarly to the low rNTP experiments but could avoid the poor signal-to-noise ratios that plague that approach. The relatively low activity of *E. coli* RNA polymerase compared to T7 RNA polymerase remained a challenge and required assay optimization. In initial work, we have demonstrated that *p*-AF aptamer ribosensors can be transcribed from a p70a *E. coli* promoter and that by adjusting assay conditions, we can recover the equilibrium signal found in T7 RNA polymerase transcription conditions at about 40% of the rate of signal generation (Supplementary Fig. S4.4b). Further work is necessary to determine whether this

platform functions properly with a slower elongation rate and whether it can be used to improve the sensitivity of existing ribosensor devices.

Furthermore, the development of an *in vitro* ribosensor platform using *E. coli* RNA polymerase would open the door to other tools for further tuning the sensitivity of the devices such as the use of *E. coli* RNA polymerase pause sites, which are not compatible with the current T7 platform. In characterized co-transcriptional kinetically-controlled RNA sensor-actuators in bacteria such as the TPP riboswitch, pause sites play a major role in providing time for ligand binding to the aptamer[120]. In the TPP riboswitch, the *ECthiC* C135 pause site has been reported to pause *E. coli* polymerase up to 10 seconds; this 30-base pause site consists of a hairpin and a 13-base unstructured sequence[40]. This mechanism has not yet been demonstrated in synthetic devices, but ongoing work in the Carothers Lab will probe the ability to reliably use this sequence for *in vitro* polymerase pausing in designed devices.

Our existing approaches- tuning binding time by elongation timers and metastable folding states- allow us to tune device t_{bind} on the order of a half second and five seconds, respectively. Transitioning to an *E. coli* RNA polymerase alone might allow us to access a few seconds of binding time, while implementing polymerase pause sites could allow us to access another 10 seconds of binding time. However, the greatest promise in engineering binding times still relies on elucidating the quantitative design rules for time encoding in metastable folding states, as kinetic traps in nature have been reported to take hundreds of seconds to resolve[121]. Metastable folding states may allow us to further improve the sensitivities of kinetic aptamer ribosensors in a length-independent manner and allow us to design devices with sensitivities that approach the aptamer K_D. Overall, by increasing the availability of the aptamer in the binding time window, we improve the range of sensitivities accessible for RNA devices constructed from

a given aptamer. This will make it increasingly possible to use aptamers for kinetic sensor applications at concentrations approaching their dissociation constant.

4.4 Supplementary Materials

Derivation of EC_{50} equation for kinetic aptamer devices

To calculate the EC_{50} of a kinetic device, we utilize the first order rate law under mass action kinetics in our system. We utilize this form of the equation as one species, in this case the small molecule, far exceeds the concentration of the ribosensor.

$$[A] = [A]_0 e^{-rt} \qquad \text{Supp. Eq. 4.1}$$

Here, [A] expresses the concentration of the unbound ribosensor and $[A]_0$ represents the initial concentration of unbound ribosensor. The first order rate law can be rearranged to give a measure of fraction of the ribosensor that is still unbound:

$$\frac{[A]}{[A]_0} = e^{-rt} \qquad \text{Supp. Eq. 4.2}$$

First, we set the fraction unbound equal to 0.5: this represents the case in which the system reaches half-maximal signal. In a kinetic system, the reaction rate between an aptamer and small molecule can be defined in terms of the association (k_{on}) and dissociation (k_{off}) rates constants; the association rate is defined by k_{on} multiplied by the concentration of the small molecule. As we are solving for the concentration at which we reach half-maximal signal, we

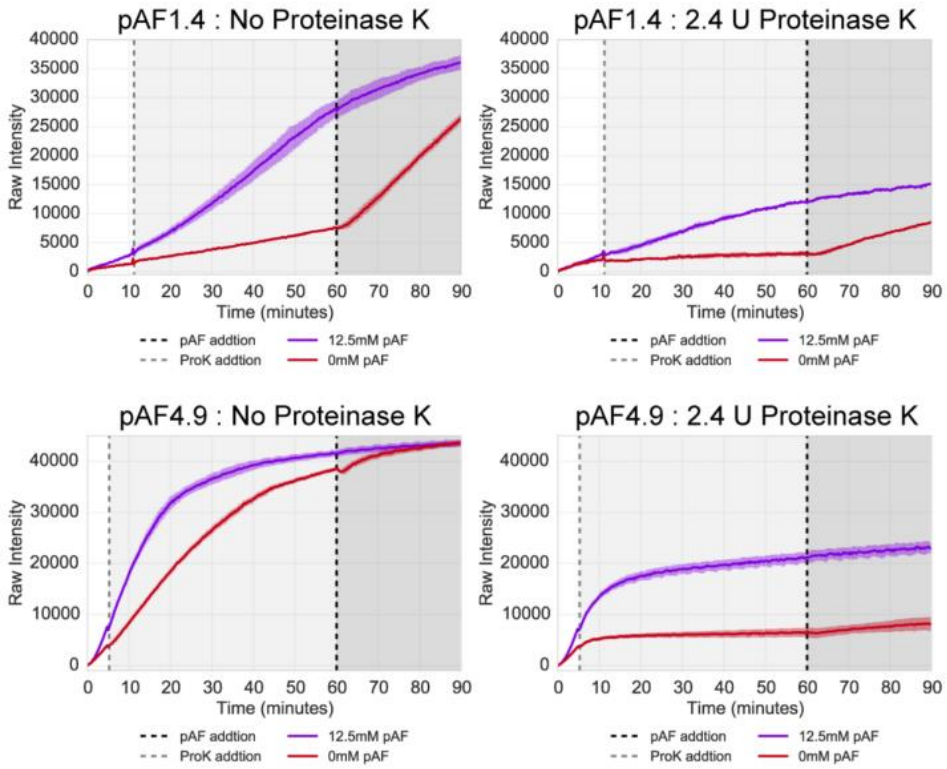
substitute the EC_{50} of the device here to define the rate $r = EC_{50}k_{on} - k_{off}$. Time here is limited to the amount of the binding window of the device, so we set the time in the equation equal to t_{bind}.

$$0.5 = e^{-(EC_{50}k_{on} - k_{off})t_{bind}} \qquad \text{Supp. Eq. 4.3}$$

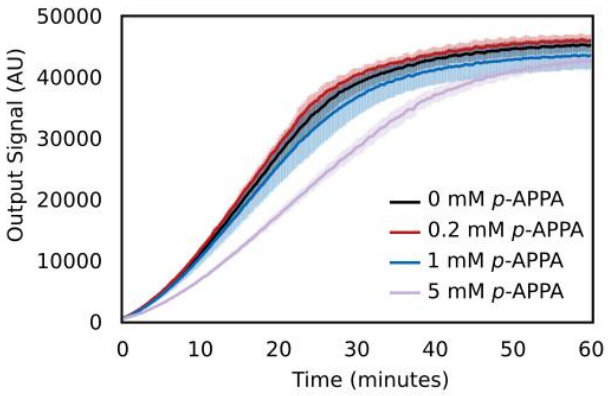
We can then rearrange this equation to solve for the EC_{50} in terms of the aptamer kinetic rate constants and t_{bind} to get Eq. 1.1.

$$EC_{50} = -\frac{ln(0.5) - k_{off} \times t_{bind}}{k_{on} \times t_{bind}} \qquad \text{Eq. 1.1}$$

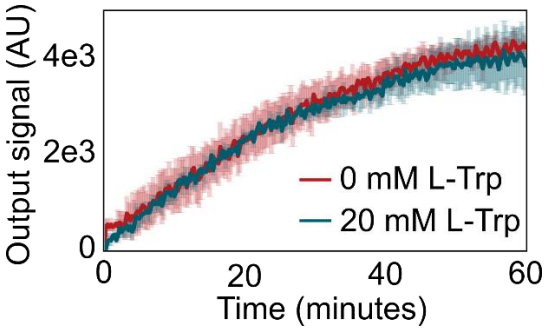
This equation can be further rearranged to solve for t_{bind} in cases where the EC_{50} has been characterized for a kinetic device.



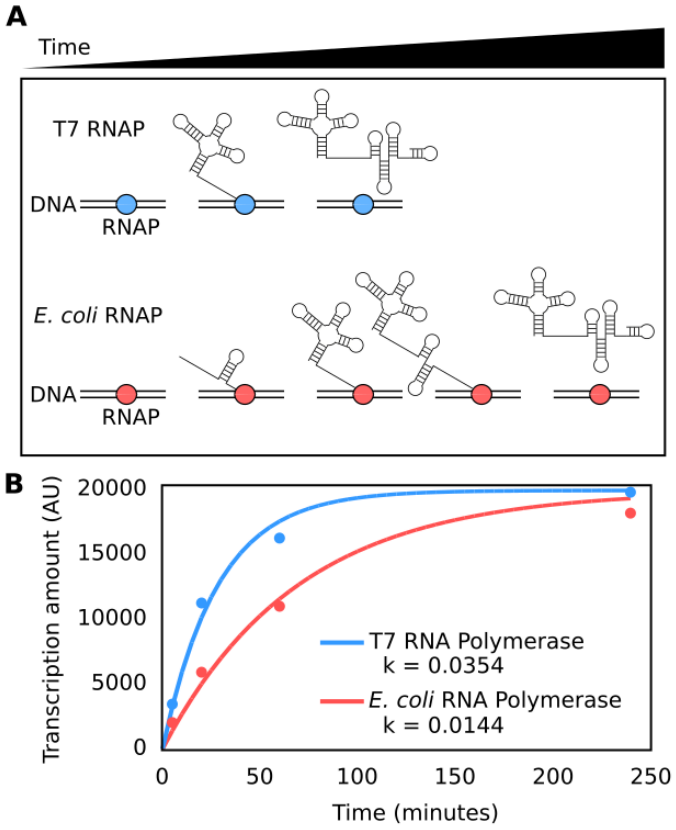
Supplementary Figure S4.1. Proteinase K assays to differentiate thermodynamic and kinetic activity in *p*-AF aptamer ribosensors. Both the pAF1.4 and pAF4.9 ribosensors were assayed in the presence and absence of Proteinase K, which degrades T7 RNA polymerase to halt transcription. In the experiments utilizing Proteinase K, the light gray dotted line marks the addition of Proteinase K, while the darker gray line indicates the addition of 12.5 mM *p*-AF to all the experiments. pAF1.4 exhibits a large ligand-dependent response after transcription has been halted with Proteinase K, while the ligand-dependent response for pAF4.9 is small.
Adapted with permission from Burke (2017).



Supplementary Figure S4.2. *In vitro* response of the pAF4.9 ribosensor to titrated levels of *p*-APPA. There is no increase in signal generation up to 1 mM *p*-APPA and a transcription defect is recorded at concentrations above 1 mM *p*-APPA, suggesting that binding of the *p*-AF aptamer to *p*-APPA is not a viable explanation for differences between HPLC titers and pAF4.9 ribosensor titers reported for quantifying *p*-AF yields from cell culture supernatants.



Supplementary Figure S4.3. Melt-and-anneal traces for the aromatic amino acid ribosensor. No difference was seen in thermodynamic signal between the melted and annealed devices at 0 mM and 20 mM tryptophan.



Supplementary Figure S4.4. Transcribing aptamer ribosensors using *E. coli* RNA polymerase could increase elongation time and device sensitivity. A. The *E. coli* RNA polymerase has a much slower characterized elongation rate than the T7 RNA polymerase; thus, effective use of the *E. coli* polymerase could increase t_{bind} and improve the sensitivities accessible to aptamer ribosensor devices. B. Through assay optimization, we have demonstrated that *E. coli* RNA polymerase can access the full overall signal level of transcription from T7 RNA polymerase and that transcript generation can reach about 40% of the rate achievable using T7 RNA polymerase in regular assay conditions.

Supplementary Table S4.1. Barrier heights for free energy barriers separating full-length secondary structure states for *p*-AF aptamer ribosensors. pAF1.4 was deliberately engineered to be under thermodynamic control. Adapted with permission from Burke (2017).

Device	Predicted free energy barrier heights (kcal/mol)				
	B1	B2	B3	B4	B5
pAF1.4	NA	NA	17.1	12.3	11.4
pAF4.5	11.7	9.8	18.3	29.0	13.4
pAF4.9	11.5	6.3	18.1	19.6	9.4
pAF4.11	12.9	9.9	18.0	17.8	16.1

Supplementary Table S4.2. Characterized EC_{50}s and designed timer lengths of pAF1.4 and three pAF4 generation devices.

Device	EC_{50}	Timer Length
pAF1.4	8 ± 2.7 mM	39 nt
pAF4.5	300 ± 50 µM	81 nt
pAF4.9	360 ± 50 µM	81 nt
pAF4.11	630 ± 180 µM	81 nt

4.5 Attribution

This chapter includes extensive work on developing the kinetic aptamer ribosensor platform by Cassandra Burke and relied on folding analyses of RNA sequences using the MFEpath algorithm developed by David Sparkman-Yager. The multi-state co-transcriptional design of these devices, including the free energy barriers that specified thermodynamic and kinetic devices and the initial idea of encoding time in metastable folding states using free energy barriers, are the work of Cassandra Burke and David Sparkman-Yager. My primary

contributions to this chapter are the experimental validations of the platform detailed here and do not include intellectual contributions to the design of this platform.

References

1. UN IPCC. *AR5 Synthesis Report: Climate Change 2014*. (2014).

2. US EPA. *Inventory of U.S. Greenhouse Gas Emissions and Sinks, 1990-2017*. (2019).

3. Wang, C. *et al.* Metabolic engineering of Escherichia coli for a-farnesene production. *Metab. Eng.* **13,** 648–655 (2011).

4. Paddon, C. J. *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* **496,** 528–532 (2010).

5. Kaur, B. & Chakraborty, D. Biotechnological and Molecular Approaches for Vanillin Production : a Review. *Appl Biochem Biotechnol* **169,** 1353–1372 (2013).

6. Goeddel, D. V *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. USA* **76,** 106–110 (1979).

7. Ajikumar, P. K. *et al.* Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science (80-.).* **330,** 70–75 (2010).

8. Kirby, J. & Keasling, J. D. Metabolic engineering of microorganisms for isoprenoid production. *Nat. Prod. Rep.* **25,** 656–661 (2008).

9. Lin, J., Wagner, J. M. & Alper, H. S. Enabling tools for high-throughput detection of metabolites : Metabolic engineering and directed evolution applications. *Biotechnol. Adv.* **35,** 950–970 (2017).

10. Keasling, J. D. Manufacturing Molecules Through Metabolic Engineering. *Science (80-.).* **330,** 1355–1359 (2010).

11. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab. Eng.* **14,** 189–195 (2012).

12. Rogers, J. K., Taylor, N. D. & Church, G. M. Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol.* **42,** 84–91 (2016).

13. Carothers, J. M. Design-driven, multi-use research agendas to enable applied synthetic biology for global health. *Syst. Synth. Biol.* **7,** 79–86 (2013).

14. Stevens, J. T. & Carothers, J. M. Designing RNA-based genetic control systems for efficient production from engineered metabolic pathways. *ACS Synth. Biol.* **4,** 107–115 (2014).

15. Blanc, V. *et al.* Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino- L -phenylalanine precursor of pristinamycin I. *Mol. Microbiol.* **23,** 191–202 (1997).

16. Carothers, J. M., Goler, J. A., Kapoor, Y., Lara, L. & Keasling, J. D. Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity. *Nucl. Acids Res.* **38,** 2736–2747 (2010).

17. Carothers, J. M., Goler, J. A., Juminaga, D. & Keasling, J. D. Model-Driven Engineering of RNA Devices to Quantitatively Program Gene Expression. *Science (80-.).* 1716–1719 (2011).

18. Burke, C. R., Yager, D. S.- & Carothers, J. M. Multi-state design of kinetically-controlled RNA aptamer ribosensors. *BioRxiv* 1–35 (2017).

19. Suvannasara, P. *et al.* Biobased Polyimides from 4 - Aminocinnamic Acid Photodimer. *Macromolecules* **47,** 1586–1593 (2014).

20. Stevens, J. T. Controlling Enzyme Expression Dynamics to Improve Production from Engineered Biosynthetic Pathways. (University of Washington, 2018).

21. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–22 (1990).

22. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (80-.).* **249,** 505–510 (1990).

23. Homann, M. & Göringer, H. U. Combinatorial selection of high affinity RNA ligands to live African trypanosomes. *Nucl. Acids Res.* **27,** 2006–2014 (1999).

24. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. a & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428,** 281–6 (2004).

25. Porter, E. B., Polaski, J. T., Morck, M. M. & Batey, R. T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat. Chem. Biol.* **13,** (2017).

26. Paige, J. S., Wu, K. Y. & Jaffrey, S. R. RNA Mimics of Green Fluorescent Protein. *Science (80-.).* **333,** (2011).

27. Mckeague, M. & Derosa, M. C. Challenges and Opportunities for Small Molecule Aptamer Development. *J. Nucleic Acids* **2012,** (2012).

28. Wilson, D. S. & Szostak, J. W. IN VITRO SELECTION OF FUNCTIONAL NUCLEIC ACIDS. *Annu. Rev. Biochem.* 611–647 (1999).

29. Pfeiffer, F. *et al.* Selection and Biosensor Application of Aptamers for Small Molecules State of the Art. *Front. Chem.* **4,** 1–21 (2016).

30. Mosing, R. K., Mendonsa, S. D. & Bowser, M. T. Capillary electrophoresis-SELEX selection of aptamers with affinity for HIV-1 reverse transcriptase. *Anal. Chem.* **77,** 6107–6112 (2005).

31. Hwang, C. & Carothers, J. M. Label-free selection of RNA aptamers for metabolic engineering. *Methods* **106,** 37–41 (2016).

32. Mckeague, M. *et al.* Analysis of In Vitro Aptamer Selection Parameters. *J. Mol. Evol.* **81,** 150–161 (2015).

33. Lou, X. *et al.* Micromagnetic selection of aptamers in microfluidic channels. *Proc. Natl. Acad. Sci. U. S. A.* **106,** (2009).

34. Szeto, K. *et al.* RAPID-SELEX for RNA Aptamers. *PLoS One* **8,** (2013).

35. Reinholt, S. J., Ozer, A., Lis, J. T. & Craighead, H. G. Highly Multiplexed RNA Aptamer Selection using a Microplate-based Microcolumn Device. *Sci. Rep.* **6,** 1–10 (2016).

36. Dausse, E. *et al.* Aptamer selection by direct microfluidic recovery and surface plasmon resonance evaluation. *Biosens. Bioelectron.* **80,** 418–425 (2016).

37. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18,** 49–60 (2005).

38. Mccown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23,** 995–1011 (2017).

39. Mishler, D. M. & Gallivan, J. P. A family of synthetic riboswitches adopts a kinetic trapping mechanism. **42,** 6753–6761 (2014).

40. Guedich, S. *et al.* Quantitative and predictive model of kinetic regulation by E . coli TPP riboswitches. *RNA Biol.* **13,** 373–390 (2016).

41. Taneda, A. MODENA : a multi-objective RNA inverse folding. *Adv. Appl. Bioinforma. Chem.* **4,** 1–12 (2011).

42. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36,** W70–4 (2008).

43. Borujeni, A. E., Mishler, D. M., Wang, J., Huso, W. & Salis, M. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucl. Acids Res.* **44,** 1–13 (2016).

44. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131,** 17303–14 (2009).

45. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Publ. Gr.* **3,** 103–113 (2011).

46. Mannironi, C., Scerch, C., Fruscoloni, P. & Tocchini-valentini, G. P. Molecular recognition of amino acids by RNA aptamers : The evolution into an L-tyrosine binder of a dopamine-binding RNA motif. *RNA* **1,** 520–527 (2000).

47. Misono, T. S. & Kumar, P. K. R. Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Anal. Biochem.* **342,** 312–317 (2005).

48. Carothers, J. M., Oestreich, S. C., Davis, J. H. & Szostak, J. W. Informational complexity and functional activity of RNA structures. *J. Am. Chem. Soc.* **63,** 57–94 (2004).

49. DeLoache, W. C. *et al.* An enzyme-coupled biosensor enables (S)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* **11,** 465–471 (2015).

50. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85,** 2149 (1963).

51. Amblard, M., Fehrentz, J.-A., Martinez, J. & Subra, G. Methods and Protocols of Modern Solid Phase Peptide Synthesis. *Mol. Biotechnol.* **33,** (2006).

52. Gao, S. *et al.* Biosensors and Bioelectronics Gonyautoxin 1 / 4 aptamers with high-af fi nity and high-speci fi city : From ef fi cient selection to aptasensor application. *Biosens. Bioelectron.* **79,** 938–944 (2016).

53. Xayaphoummine, a., Bucher, T. & Isambert, H. Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res.* **33,** 605–610 (2005).

54. Kellenberger, C. A., Wilson, S. C., Sales-lee, J. & Hammond, M. C. RNA-Based Fluorescent Biosensors for Live Cell Imaging of Second Messengers Cyclic di-GMP and Cyclic AMP-GMP. *J. Am. Chem. Soc.* **135,** 4906–4909 (2013).

55. Abatemarco, J. *et al.* RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nat. Commun.* **8,** (2017).

56. Andrews, P. R., Craik, D. J. & Martin, J. L. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* **27,** 1648–1657 (1984).

57. Baum, B. *et al.* Non-additivity of Functional Group Contributions in Protein – Ligand Binding : A Comprehensive Study by Crystallography and Isothermal Titration Calorimetry. *J. Mol. Biol.* **397,** 1042–1054 (2010).

58. Lydersen, A. L. *Estimation of Critical Properties of Organic Compounds by the Method of Group Contributions*. *University of Wisconsin* (1955).

59. Joback, K. G. & Reid, R. C. Estimation of pure-component properties from group-contributions. *Chem. Eng. Commun.* **57,** 233–243 (1987).

60. Mavrovouniotis, M. L. Group contributions for estimating standard Gibbs free energies of formation of biochemical compounds in aqueous solution. *Biotechnol Bioeng* **36,** 1070–1082 (1990).

61. Jankowski, M. D., Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Group Contribution Method for Thermodynamic Analysis of Complex Metabolic Networks. *Biophys. J.* **95,** 1487–1499 (2008).

62. Majerfeld, I. & Yarus, M. Isoleucine : RNA sites with associated coding sequences. *RNA* **4,** 471–478 (1998).

63. Majerfeld, I. & Yarus, M. An RNA pocket for an aliphatic hydrophobe. *Nat. Struct. Mol. Biol.* **1,** 287–292 (1994).

64. Illangasekare, M. & Yarus, M. Phenylalanine-Binding RNAs and Genetic Code Evolution. *J Mol Evol* **54,** 298–311 (2002).

65. Majerfeld, I., Puthenvedu, D. & Yarus, M. RNA Affinity for Molecular L-Histidine ; Genetic Code Origins. *J Mol Evol* **61,** 226–235 (2005).

66. Majerfeld, I. & Yarus, M. A diminutive and specific RNA binding site for L-tryptophan. *Nucl. Acids Res.* **33,** 5482–5493 (2005).

67. Davis, J. H. & Szostak, J. W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. **99,** 8–13 (2002).

68. Geiger, A. *et al.* RNA aptamers that bind L -arginine with sub-micromolar dissociation constants and high enantioselectivity. **24,** 1029–1036 (1996).

69. Famulok, M. Molecular Recognition of Amino Acids by RNA- Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *J. Am. Chem. Soc.* **116,** 1698–1706 (1994).

70. Kiga, D., Futamura, Y., Sakamoto, K. & Yokoyama, S. An RNA aptamer to the xanthine / guanine base with a distinctive mode of purine recognition. **26,** 1755–1760 (1998).

71. Koizumi, M., Breaker, R. R., Biology, D. V, V, Y. U. & Box, P. O. Molecular Recognition of cAMP by an RNA Aptamer. *Biochemistry* **39,** 8983–8992 (2000).

72. Lauhon, C. T. & Szostak, J. W. RNA Aptamers That Bind Flavin and Nicotinamide Redox Cofactors. *J. Am. Chem. Soc.* **117,** 1246–1257 (1995).

73. Meli, M., Vergne, J. & Maurel, M. Adenine-Aptamer Complexes. *J Biol Chem* **277,** 2104–2111 (2002).

74. Burgstaller, P. & Famulok, M. Isolation of RNA Aptamers for Biological Cofactors. *Angew. Chem., Int. Ed.* **33,** 1084–1087 (1994).

75. Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. *Nature* **364,** 550–553 (1993).

76. Wang, Y. & Rando, R. R. Specific binding of aminoglycoside antibiotics to RNA. *Chem. Biol.* **2,** 281–290 (1995).

77. Win, M. N., Klein, J. S. & Smolke, C. D. Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucl. Acids Res.* **34,** 5670–5682 (2006).

78. Mannironi, C., Nardo, A. Di & Fruscoloni, P. In Vitro Selection of Dopamine RNA Ligands. *Biochemistry* **2960,** 9726–9734 (1997).

79. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-Resolution Molecular Discrimination by RNA. *Science (80-.).* **263,** 1425–1429 (1994).

80. Grate, D. & Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc Natl Acad Sci USA* **96,** 6131–6136 (1999).

81. Haller, A. & Sarnow, P. In vitro selection of a 7-methyl-guanosine binding RNA that inhibits translation of capped mRNA molecules. *Proc Natl Acad Sci USA* **94,** 8521–8526 (1997).

82. Bala, J., Bhaskar, A., Varshney, A., Singh, A. K. & Yadava, P. In vitro selected RNA aptamer recognizing glutathione induces ROS mediated apoptosis in the human breast cancer cell line MCF 7. **6286,** (2011).

83. Lato, S. M., Boles, A. R. & Ellington, A. D. In vitro selection of RNA lectins: using combinatorial chemistry to interpret ribozyme evolution. *Chem. Biol.* **2,** 291–303 (1995).

84. Gebhardt, K., Shokraei, A., Babaie, E. & Lindqvist, B. H. RNA Aptamers to S-Adenosylhomocysteine: Kinetic Properties, Divalent Cation Dependency , and Comparison with Anti-S-adenosylhomocysteine Antibody. *Biochemistry* **39,** 7255–7265 (2000).

85. Horii, K. *et al.* Development of a Sphingosylphosphorylcholine Detection System Using RNA Aptamers. *Molecules* **15,** 5742–5755 (2010).

86. Brockstedt, U., Uzarowska, A., Montpetit, A., Pfau, W. & Labuda, D. In vitro evolution of RNA aptamers recognizing carcinogenic aromatic amines. *Biochem. Biophys. Res. Commun.* **313,** 1004–1008 (2004).

87. Wilson, C. & Jack, B. In vitro evolution of a self-alkylating ribozyme. *Nature* **374,** 777–782 (1995).

88. Berens, C., Thain, A. & Schroeder, R. A Tetracycline-binding RNA Aptamer. *Bioorg. Med. Chem.* **9,** 2549–2556 (2001).

89. Hermann, T. & Patel, D. J. Adaptive Recognition by Nucleic Acid Aptamers. *Science (80-.).* **287,** 820–826 (2000).

90. Kass, R. & Raftery, A. Bayes Factors. *J. Am. Stat. Assoc.* **90,** 773–795 (1995).

91. Xiu, Y., Jang, S., Jones, J. A., Zill, N. A. & Linhardt, R. J. Naringenin-Responsive Riboswitch-Based Fluorescent Biosensor Module for Escherichia coli Co-Cultures. *Biotechnol. Bioeng.* **9999,** 1–10 (2017).

92. J, X., Carrocci, T. & Hoskins, A. Evolution and characterization of a benzylguanine-binding RNA aptamer. *Chem. Commun.* **52,** 549–552 (2016).

93. Mckeague, M. *et al.* Comprehensive Analytical Comparison of Strategies Used for Small Molecule Aptamer Evaluation. *ACS Anal. Ch* **87,** 8608–8612 (2015).

94. Brown, R. F., Andrews, C. T. & Elcock, A. H. Stacking Free Energies of All DNA and RNA Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised AMBER Parameters and Compared with Experiment. *J Chem Theory Comput* **11,** 2315–23828 (2015).

95. Perola, E. & Charifson, P. S. Conformational Analysis of Drug-Like Molecules Bound to

Proteins : An Extensive Study of Ligand Reorganization upon Binding. *J. Med. Chem.* **47,** 2499–2510 (2004).

96. Nicklaus, M. C., Wang, S., Driscoll, J. S. & Milne, G. W. A. Conformational Changes of Small Molecules Binding to Proteins. *Bioorg. Med. Chem.* **3,** 411–428 (1995).

97. Mobley, D. L. & Dill, K. A. Review Binding of Small-Molecule Ligands to Proteins : ‘“ What You See ”’ Is Not Always ‘“ What You Get ”.’ *Structure* **17,** 489–498 (2009).

98. Huang, L., Yang, X., Qi, C., Niu, X. & Zhao, C. A label-free electrochemical biosensor based on a DNA aptamer against codeine. *Anal. Chim. Acta* **787,** 203–210 (2013).

99. Harada, K. & Frankel, A. D. Identification of two novel arginine binding DNAs. *EMBO J.* **14,** 5798–5811 (1995).

100. Niazi, J. H., Lee, S. J. & Gu, M. B. Single-stranded DNA aptamers specific for antibiotics tetracyclines. *Bioorg. Med. Chem.* **16,** 7245–7253 (2008).

101. Kato, T., Takemura, T., Yano, K., Ikebukuro, K. & Karube, I. In vitro selection of DNA aptamers which bind to cholic acid. *Biochim. Biophys. Acta* **1493,** 12–18 (2000).

102. Kim, Y. S., Hyun, C. J., Kim, I. A. & Gu, M. B. Isolation and characterization of enantioselective DNA aptamers for ibuprofen. *Bioorg. Med. Chem.* **18,** 3467–3473 (2010).

103. Yang, Q., Goldstein, I. J., Mei, H.-Y. & Engelke, D. R. DNA ligands that bind tightly and selectively to cellobiose. *Proc Natl Acad Sci USA* **95,** 5462–5467 (1998).

104. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc. Natl. Acad. Sci.* **104,** 14283–14288 (2007).

105. Lee, S. Y. *et al.* A comprehensive metabolic map for production of bio-based chemicals. *Nat. Catal.* **2,** 18–33 (2019).

106. Halls, C. & Yu, O. Potential for metabolic engineering of resveratrol biosynthesis. *Trends Biotechnol* **26,** 77–81 (2008).

107. Jiang, H., Wood, K. V & Morgan, J. A. Metabolic Engineering of the Phenylpropanoid

Pathway in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **71,** 2962–2969 (2005).

108. Thodey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platorm for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10,** 837–844 (2014).

109. Levesque, D., Beaudoin, J., Roy, S. & Perreault, J.-P. In vitro selection and characterization of RNA aptamers binding thyroxine hormone. *Biochem J* **403,** 129–138 (2007).

110. Chang, A. L., Mckeague, M., Liang, J. C. & Smolke, C. D. Kinetic and Equilibrium Binding Characterization of Aptamers to Small Molecules using a Label-Free, Sensitive, and Scalable Platform. *ACS Anal. Ch* **86,** 3273–3278 (2014).

111. Shoji, A., Kuwahara, M., Ozaki, H. & Sawai, H. Modified DNA Aptamer That Binds the (R)-Isomer of a Thalidomide Derivative with High Enantioselectivity. *J. Am. Chem. Soc.* **129,** 1456–1464 (2007).

112. Karlsen, K. K. & Wengel, J. Locked Nucleic Acid and Aptamers. *Nucleic Acids Ther.* **22,** 366–370 (2012).

113. Rangel, A. E., Chen, Z., Ayele, T. M. & Heemstra, J. M. In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucl. Acids Res.* **46,** 8057–8068 (2018).

114. Uhlenbeck, O. C. Keeping RNA happy. *RNA* **1,** 4–6 (1995).

115. Treiber, D. K. & Williamson, J. R. Exposing the kinetic traps in RNA folding. *Curr. Opin. Struct. Biol.* **9,** 339–345 (1999).

116. Shin, J. & Noireaux, V. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* **1,** 29–41 (2012).

117. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J. Bacteriol.* **174,** 619–622 (1992).

118. Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E. & Hoggett, J. G. Promoter Binding , Initiation , and Elongation By Bacteriophage T7. *J. Biol. Chem.* **279,** 3239–3244 (2004).

119. Tolic, S. F., Engh, A. M., Landick, R. & Gelles, J. Diversity in the Rates of Transcript Elongation by Single RNA. *J. Biol. Chem.* **279,** 3292–3299 (2004).

120. Chauvier, A. *et al.* Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat Commun* **8,** (2017).

121. Gultyaev, A. P., Franch, T. & Gerdes, K. Programmed Cell Death by hok / sok of Plasmid R1 : Coupled Nucleotide Covariations Reveal a Phylogenetically Conserved Folding Pathway in the hok Family of mRNAs. *J. Mol. Biol.* **273,** 26–37 (1997).