# Expanding the Toolkit for Synthetic Biology: Frameworks for Native-like Non-natural Gene Circuits

Thesis by

Emmanuel Lorenzo Cornejo de los Santos

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy in Bioengineering


California Institute of Technology

Pasadena, California

2015

(Defended May 11, 2015)

# Acknowledgements

First and foremost, I'd like to thank God for all the opportunities He's given me throughout the years. I'm able to do things I'm very interested in and passionate about and for that I'm forever grateful. All my work is for His greater glory.

I'd like to thank my family–Manny, Monette, Jem, and Mariel–for their unconditional love and support through good times and bad. I'd also like to thank in a special way Ninang Arlene, Rafa, and Anton for providing me a home and a second family here in the States, as well as my other relatives for their support and prayers.

I'd like to thank the my teachers, professors, and mentors in the schools I've attended. Philippine Science High School opened up a lot of opportunities for me and fostered my curiosity for math and science. I'd like to thank Ma'am Vea (Espiritu) in particular for starting my interest in molecular biology. I'm only one of your many former students that are pursuing biology-related fields because of your biology class. My professors at MIT continued to foster and support my evolving interests. Linda, Doug, Scott, and Steve were among the faculty in the new undergraduate Biological Engineering major while I was there whom I had the pleasure of interacting with. They really fostered an open environment and listened to our needs in the development of the new major. Amy took a chance on me when I had no research experience and was really a mentor to me and got me interested in proteins and computational protein design. She's helped me even as I moved to Caltech. I'd like to thank Drew for first getting me interested in Synthetic Biology. Nora and Shaun helped me immensely when I had no idea what I was doing in my days in lab and programming. I had an immense amount of faculty and graduate student support during my time here in Caltech. I'd like to thank my advisers, Richard and Steve, for all their advice and mentorship throughout my

PhD. I was very lucky to have two advisers that had different strengths whom I could draw on for guidance and advice. Dave and Michael were gracious enough to be on my thesis committee and offer their advice when I asked for it. The Mayo and Murray labs also provided a good environment for research and I was able to talk to a lot of people for mentorship and advice. I'd like to thank specifically the people I've worked with over the years in both labs, including Victoria, Joe, Tim, and Bernardo.

Finally, I'd like to thank all my friends both near and far, in Caltech, church, college, high school, and home for listening to me when I needed it, and all of the talks, trips, movies, food adventures, boardgames, and whatnot. You truly helped me keep my sanity throughout this entire process and made the whole experience a very positive and enriching one.

# Abstract

Synthetic biology combines biological parts from different sources in order to engineer non-native, functional systems. While there is a lot of potential for synthetic biology to revolutionize processes, such as the production of pharmaceuticals, engineering synthetic systems has been challenging. It is oftentimes necessary to explore a large design space to balance the levels of interacting components in the circuit. There are also times where it is desirable to incorporate enzymes that have non-biological functions into a synthetic circuit. Tuning the levels of different components, however, is often restricted to a fixed operating point, and this makes synthetic systems sensitive to changes in the environment. Natural systems are able to respond dynamically to a changing environment by obtaining information relevant to the function of the circuit. This work addresses these problems by establishing frameworks and mechanisms that allow synthetic circuits to communicate with the environment, maintain fixed ratios between components, and potentially add new parts that are outside the realm of current biological function. These frameworks provide a way for synthetic circuits to behave more like natural circuits by enabling a dynamic response, and provide a systematic and rational way to search design space to an experimentally tractable size where likely solutions exist. We hope that the contributions described below will aid in allowing synthetic biology to realize its potential.

# Contents

# Chapter 1

# Introduction

Synthetic biology aims to take advantage of biological processes to engineer systems and devices with novel and useful functions. One promising application is the engineering of organisms to produce a wide range of products from fuels to cosmetics to pharmaceuticals. Perhaps the most successful example of the potential of synthetic biology is the story of artemisinin, a potent anti-malarial drug. Researchers engineered yeast to produce artemisinic acid, a precursor of artemisinin, in a practical and scalable way [1]. This method of producing artemisinic acid was used by Sanofi Aventis as part of a fully implemented industrial process to produce the drug semi-synthetically [2]. The process involves the fermentation of artemisinic acid by genetically modified yeast developed by Keasling and Amyris, followed by an efficient synthetic process to convert artemisinic acid to artemisinin. In 2014, sixty tons of artemisinin were produced in this way, supplementing the highly volatile supply which involves the drugs extraction from the leaves of *Artemisia annua* [2]. While highlighting the promise of synthetic biology as a transformative technology, the artemisinin story is also informative to the areas where research and development are needed in order for synthetic biology to realize its full potential.

Developing an industrially feasible process for the production of artemisinin was a massive effort that took more than ten years. The first report on genetically engineered yeast for the production of artemisinic acid came out in 2006 [1]. This work leveraged even older work done on engineering a mevalonate pathway in *E. coli*, published in 2003, for the production of terpenoids, the class of molecules to which artemisinin belongs [3]. Even after the successful development of a yeast strain

that produced artemisinic acid, seven years of further strain optimization to optimize gene levels, induction, and the addition of new genes were required to engineer yeast that produced commercially relevant quantities of the drug precursor [4].

Even at its current state, the price of artemisinin derived from the process realized by synthetic biology is still comparable to the product derived from agriculture [5]. It serves as another source of artemisinin and reduces the price volatility, but the promise of cheap artemisinin from the synthetic biology derived process to meet the world's global supply has not been realized. The leaders of the artemisinin project have been criticized for overselling and under-delivering the technology. They have also been criticized for oversimplifying the global health problem that also involves multiple factors outside of science and technology, such as the socioeconomic effects of cheap artemisinin of *Artemisia* farmers and the economics of requiring combination therapies of anti-malarial drugs in the developing world.

Further development of the technology involved in the process of developing a synthetic biology route to produce artemisinin can address some of these concerns. Demonstrating the ability to use synthetic biology to tackle the development of other small molecules in an industrially relevant scale can show that artemisinin is not a single success story, but a way forward, where the molecules targeted are determined by considering the different complex factors involved. Having other molecules to point to can also help reduce the tendency to oversell the impact of artemisinin as we can begin to refer to a process to obtain target molecules of interest using synthetic biology instead of just having a single molecule as an example. While some of the technology and lessons learned from the development of the artemisinin like the bioinformatics, genetic engineering, and metabolic modeling techniques can readily be applied in future efforts. It is currently unclear how much effort would be necessary to develop strains that produce industrially relevant quantities of compounds that are chemically different and thus are at a different metabolic space from isoprenoids.

This is particularly relevant in the discovery and development of new antibiotics to combat multi-drug resistant bacteria. Since bacteria have developed mechanisms of resistance to the commonly used chemical scaffolds in existing antibiotics, it is possible that new antibiotics discovered that we

would want to mass produce are based on new chemical scaffolds whose production is mediated by genes and biosynthetic pathways that are not well characterized [6, 7].

To this end, we need to develop a set of readily generalizable frameworks and tools that accelerate the process from a target molecule to a fully realized industrial process. This thesis describes our efforts towards this. Chapter 2 describes a way to give the cell information about the circuit it is executing. This is done by engineering a transcription factor to sense a small molecule that can be related to the particular pathway of interest. Chapter 3 addresses the problem of having genes from different sources causing imbalances in the gene network by regulating the ratios of different genes in a dynamic and tunable way. To achieve this, we use synthetic scaffolds coupled with a histidine kinase-response regulator pair that is independent of the gene circuit. Chapter 4, provides a framework for synthetic metabolic pathways to reach new areas of chemical space by designing enzymes that have novel function, and translating desired reactions into a language that cells can process. We do this by systematically searching the space of possibilities in a systematic way, narrowing down a list of potential solutions to a few proteins that we can test experimentally. Appendix A discusses our work to engineer a mechanism for general allosteric control in proteins through the insertion of a domain that changes conformation upon phosphorylation. This attempt was hampered by our inability to find a reliable high-throughput assay.

The work here contains two major themes: first, the design of tools that allow dynamical response in synthetic circuits, and second, the development of frameworks to systematically and rationally search a large design space.

Cells are able to dynamically respond to their environment. They execute genetic programs that are able to sense the available resources and molecules around them and alter the specific proteins they produce and their quantity based on their surroundings. This is mediated by a set of sensors and regulators whose exact mechanisms and functions we are still discovering. These components form the basis for the genetic controllers that execute programs and allow the cells dynamic response. Synthetic biologists have used these components to create synthetic circuits that perform operations needed for dynamically responsive genetic controllers. While these circuits are useful, as they are

able to display the function and logic necessary for a controller, such as an "and" operation or an "on" signal, their use is often limited to a proof-of-concept or model examples, since they respond to signals related to the original function of the circuit they are taken from [8]. Cells will have the machinery in place to know when they have too much or too little of a metabolite they encounter, or too much of a specific protein in the specific context that the circuit evolved. However, since synthetic circuits implement non-native programs in their hosts, the cells have no such machinery in place when we want them to respond to this foreign genetic program. The cell does not "know" what the program is trying to accomplish. This is problematic in the case when the circuit you are trying to implement contains toxic intermediates, or requires the careful balancing of different proteins in the circuit. To enable native-like behavior in synthetic circuits, we need to establish a framework that facilitates communication between the cell and the new gene circuit. The ability to track and obtain information about a circuit you wish to control is essential for this, as it allows us to "teach" the cell about the program we want it to execute and enables it to alter specific parts of the circuit with changing environmental contions. The benefits of a dynamical response in a pathway enabled by molecules tied to the actual pathway have been demonstrated in the increased yields in efforts to produce fatty acid derived chemicals and fuels in *E. coli* [9].

We aim to address this by developing mechanisms in which cells can track levels of different small molecules (Chapter 2), or proteins (Chapter 3). Chapter 2 discusses a framework that can enable the development of sensors that are able to detect levels of specific small molecules. This enables one to turn on and off specific genes based on their necessity. Chapter 3 covers a complementary sensing mechanism that is able to track the level of a specified protein and match it to a reference level of another specified protein. These two systems could be used in conjunction to allow the cell to efficiently allocate its resources to maximize the production of non-native products, as it is able to respond to levels of small molecules and proteins that are specifically part of that pathway as if it were a native program in the cell. The small molecule sensors in Chapter 2 can set the overall protein levels for the circuit, while a scaffold based mechanism in Chapter 3 would ensure that the ratios of enzymes in the circuit are at the optimum level.

Part of the strain engineering required to increase the yield of artemisinin to industrially relevant quantities involved modifying promoters to respond to different signals, and using different sets of promoters to match gene expression levels. Since these were not tied to genes and molecules in the artemisinin pathway, they are optimized for specific conditions and cannot respond dynamically to changes in the environment that directly affect the target pathway.

In the process of designing tools that had specific desired function, we encountered the problem of searching through large design spaces. The number of potential mechanisms and solutions that could address a problem was large. It was not feasible or tractable to explore the entire design space. As such, it was necessary to develop frameworks that allowed us to systematically and rationally reduce the design space we wanted to explore to a point where it was experimentally tractable. Chapter 2 covers the framework we used in systematically reducing the design space in order to obtain the sensors to the small molecules we desired. Chapter 4 discusses the systematic exploration of design space in our efforts to engineer *de novo* an enzymatic function into an inactive scaffold. In both of these cases, the systematic reduction of the area of design space to explore was necessary for the work to proceed. Neither of the projects would have been attempted had such a framework not existed since the design space would simply have been too large. Appendix A describes our attempt in engineering modular allosteric feedback, and also provides an example of the challenges when no such framework to systematically explore a large design space is in place.

Advances in DNA sequencing and synthesis make systematic methods to explore design space become even more important in synthetic biology. The large amount of data that is being collected gives us a large repository of sequences that can be used both as functional components for other synthetic circuits or the foundations of gene circuits that produce molecules with interesting properties. The reduced cost of DNA synthesis allows us to test these components even if their source is unculturable or unknown. However this can be a double-edged sword. It is impossible to characterize and test all of the functional sequences we obtain given the limited experimental resources. A framework for the systematic exploration of these sequences to determine areas that warrant further exploration is necessary. Intelligent exploration and characterization of the information and parts

we have will ensure that we are maximizing the resources we have and realizing the potential we have from these technological advances.

The different tools and frameworks discussed in this work expand the toolkit that is available for use in the rational design of genetic programs into cells and are a step in fully realizing the potential of synthetic biology.

# Chapter 2

# Engineering Transcriptional Regulator Effector Specificity using Computational Design and *In Vitro* Rapid Prototyping: Developing a Vanillin Sensor

## 2.1 Introduction

The pursuit of engineering cells that contain circuits and novel metabolic pathways of increasing complexity and robustness in synthetic biology will require engineering new regulatory tools. The utility of a synthetic genetic circuits for real world applications is dependent on the ability to effectively trigger the circuit. While we can control the expression of target genes with transcriptional regulators, triggers for these transcriptional regulators are limited to a small number of molecules and other inputs (e.g., light) [10]. As a consequence, most synthetic circuits right now are limited to proof-of-principle demonstrations without being extendable to real world applications. Feedback control based on relevant molecules, including toxic intermediates and environmental signals, would enable genetic circuits to react appropriately to changing conditions. This requires us to be able to transmit the levels of the relevant molecules to existing transcriptional control machinery.

This chapter develops a framework to use a combination of sequence generation by computational

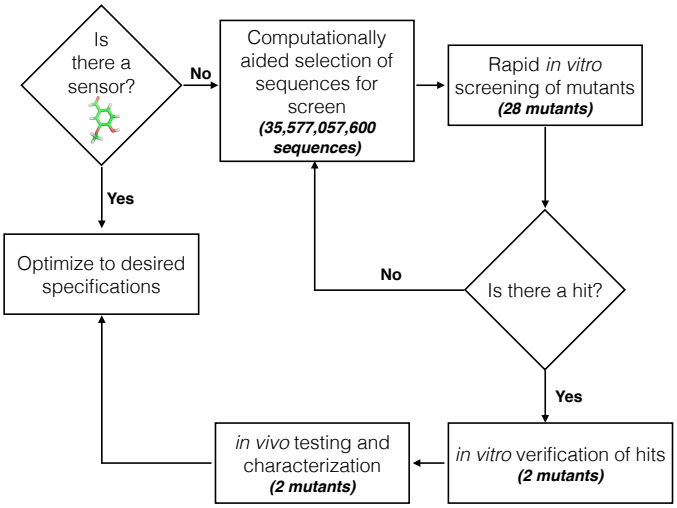



Figure 2.1: **Workflow for generating novel sensors.** The *in vitro* TX-TL platform allows for the rapid screening of sequences selected with the help of computational protein design. Hits from the *in vitro* screen are then verified by further *in vitro* testing. *In vivo* testing and characterization can then be performed to see if they meet the desired specifications. Further refinement of the hits through directed evolution or further computational design can be performed until necessary specifications are achieved. Numbers in parentheses are the number of sequences considered by the computational algorithm, or the number of mutants assayed at the specified step for vanillin.

protein design (CPD) and rapid prototyping using a cell-free transcription-translation (TX-TL) system to switch effector specificity of existing transcriptional regulators to respond to targeted small molecules of interest (Figure 2.1).

The small molecule we chose to target was vanillin, a byproduct from the lignin degradation done in the acid pre-treatment of ligno-cellulosic biomass for ethanol fermentation. Lignin crosslinks cellulose and hemi-cellulose, the main carbon source in biofuel production, making them difficult to access. However, the resulting byproducts from lignin are growth inhibitors and are difficult to separate [11]. In developing a sensor for vanillin or similar lignin degradation byproducts, we can design a synthetic feedback loop for stress response to these growth inhibitors. This can be done by controlling the expression of genes to mitigate vanillin toxicity such as through efflux pumps, or by enzymes which convert vanillin into a less toxic molecule. The expression of some of these genes could be metabolically expensive for the cell and their constitutive expression could have a negative effect on biofuel yield and cell growth.

Transcription factor engineering has primarily focused on modifying DNA-binding specificity [12]

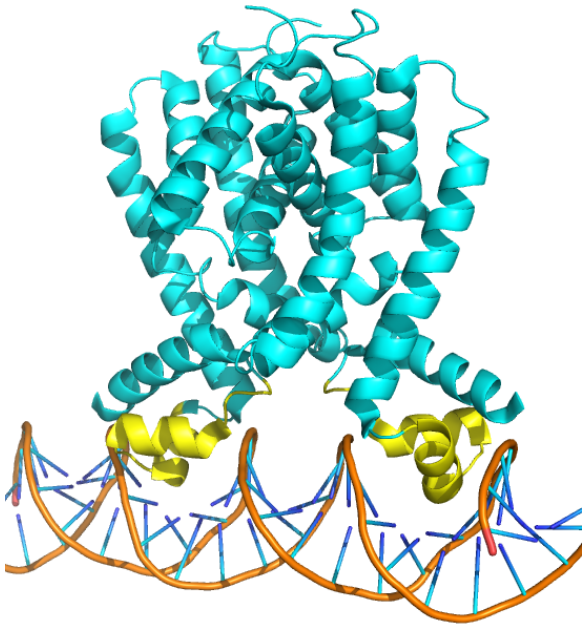


Figure 2.2: **QacR bound to DNA.** Cartoon representation of a crystal structure of qacR (PDB ID: 1JT0), a tetR-family repressor, bound to DNA. The helix-turn-helix domain is colored in yellow and the ligand-binding domain is colored in cyan. QacR binds to DNA as a dimer.

or altering the effector specificity to similar molecules from the original effector [13]. To the best of our knowledge, this is the first instance of effector specificity engineering where the target small molecule effector was chosen independent of the transcriptional repressor and was an unrelated, dissimilar small molecule to its natural effectors. A version of this chapter has been submitted to *ACS Synthetic Biology* and is available online in the *bioRxiv* [14]. The work described here was a collaboration between myself and Joseph Meyerowitz. I came up with the concept of using computational protein design to switch the effector specificity of a transcription factor. Joseph suggested vanillin and performed the preliminary *in vitro* screens. I performed the further *in vitro* characterization and the *in vivo* tests.

## 2.2 Background

The tetR family is a large family of transcriptional regulators found in bacteria. They are named after the tetR repressor, which controls the expression of tetA, an efflux pump for tetracycline [15]. They contain two domains: a helical-bundle ligand-binding domain and a helix-turn-helix DNA-

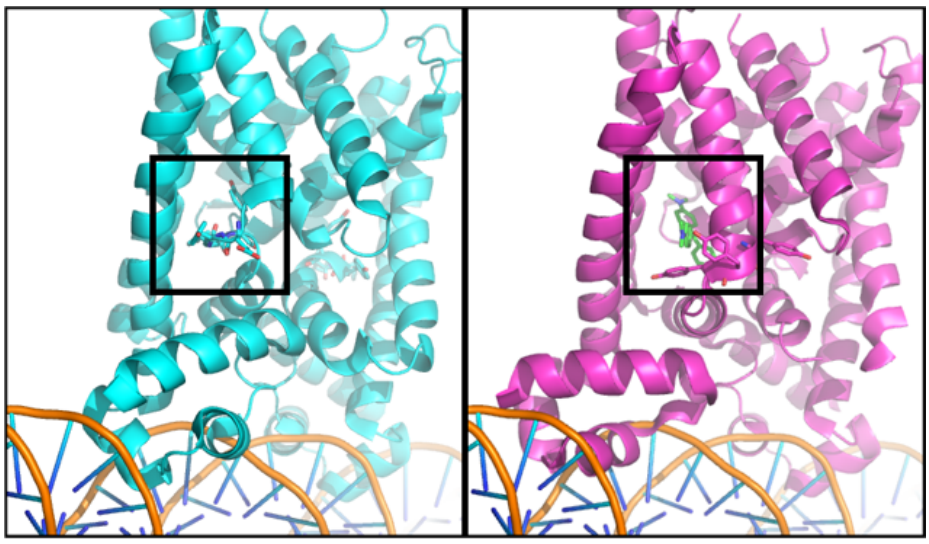
binding domain. In the absence of their inducing molecule, tetR repressors bind to DNA, preventing the transcription of downstream genes (Figure 2.2). Inducer binding to the ligand-binding domain causes a conformational change in the DNA binding domain that causes dissociation from the DNA, allowing transcription of downstream genes. The tetR transcriptional regulation machinery has been used in the design of synthetic circuits, including the repressilator [16] and the toggle switch [17].

QacR is a tetR-family repressor found in *S. aureus* that controls the transcription of qacA, an efflux pump that confers resistance to a large number of quaternary anionic compounds. The protein has been studied because it is induced by a broad range of structurally dissimilar compounds [18]. Structural examination of qacR in complex with different small molecules has shown that qacR has two different binding regions inside a large binding pocket. While qacR has multiple binding modes for various inducers, in all cases for which there are structures, binding of the inducer causes a tyrosine expulsion that moves one of the helices and alters the conformation of the DNA binding domain, rendering qacR unable to bind DNA [19, 20, 21]. Crystal structures of inducer-bound forms of qacR and the qacR-DNA complex coupled with a definitive structural mechanism for qacR induction make it the ideal starting point for CPD of new transcriptional regulators. In this work, we describe our efforts to apply our framework to engineer qacR to sense vanillin, a phenolic growth inhibitor that is a byproduct of lignin degradation performed during the processing of biomass into intermediate feedstock in biofuel production [22].

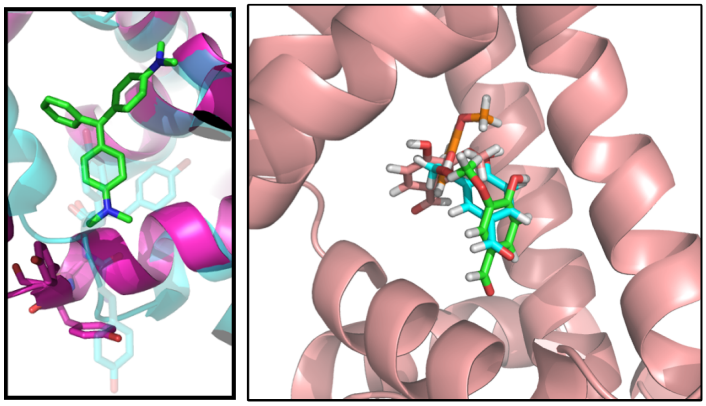
## 2.3 Results and Discussion

### 2.3.1 Computationally Aided Selection of QacR Mutant Sequences

We created a computational model of vanillin to place into a crystal structure of QacR (PDB ID: 1JTO). A computational protein design algorithm was used to find potential vanillin binding sites close to the location of the tyrosine expulsion in the binding pocket of qacR (Figure 2.3A-B) while being in the proximity of amino acid positions that allowed for favorable pi-stacking and hydrogen bonding interactions. We used targeted ligand placement [23] to find potential binding



(a)



(b) (c)

Figure 2.3: **Computationally aided selection of qacR mutants. (a)** PDB structures of the non-ligand bound (cyan, PDB ID: 1JTO) and ligand bound (magenta, PDB ID: 3BQZ) conformations of qacR. A conformational shift in the binding pocket occurs upon entry of the small molecule, causing the protein to dissociate from DNA. **(b)** A closer look at the binding pocket of qacR: the binding of the ligand in green causes the displacement of three tyrosine residues, shown as sticks in cyan and magenta. **(c)** Computational model for potential vanillin binding sites. Vanillin is shown as a different color in each of the four sites. A protein design algorithm was asked to suggest mutations for amino acids close to the potential binding sites to support the placement of vanillin in these sites.

Table 2.1: **Initial qacR mutants chosen through computationally guided design.** List of the amino acid mutations by position of the initial set of qacR mutants selected. Low energy sequences from different optimization runs were analyzed and a set of 10 mutants was selected for testing.

| **Protein** | **Amino Acid Position** | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 54 | 57 | 58 | 61 | 86 | 89 | 90 | 93 | 96 | 99 | 102 | 116 | 119 | 120 | 126 | 154 | 157 | 161 |
| **qacR-wt** | F | L | E | E | W | S | T | E | Y | Q | I | F | M | L | E | A | N | N | T |
| mutant1 | A | W | Q | L | Y | S | T | Q | Y | M | Q | S | Q | Y | Q | A | M | Q | M |
| mutant2 | A | W | Q | L | Y | S | T | Q | Y | Q | Q | F | Q | Y | Q | A | M | L | M |
| mutant3 | A | W | Q | L | Y | S | T | Q | Y | Q | I | S | Q | Y | Q | A | M | L | M |
| mutant4 | A | W | Q | L | Y | S | T | Q | Y | M | Q | Q | M | Y | Q | A | M | Q | M |
| mutant5 | A | W | E | E | Y | S | T | Q | Y | M | Q | S | Q | Y | Q | A | N | N | T |
| mutant6 | A | W | E | E | Y | S | T | Q | Y | Q | Q | F | Q | Y | Q | A | N | N | T |
| mutant7 | A | W | E | E | Y | S | T | Q | Y | Q | I | S | Q | Y | Q | A | N | N | T |
| mutant8 | A | W | E | E | Y | S | T | Q | Y | M | Q | Q | M | Y | Q | A | N | N | T |
| mutant9 | A | W | Q | L | W | S | T | Q | Y | M | Q | S | Q | Y | Q | A | M | Q | M |
| mutant10 | A | W | E | E | W | S | T | Q | Y | Q | I | F | M | Y | Q | A | N | N | T |

positions for vanillin by defining an idealized binding site for the molecule. The algorithm yielded four potential binding positions for vanillin (Figure 2.3C). Computational protein sequence design was then used to select amino acid residues at positions around the potential vanillin binding sites. In order to minimize the possibility of steric clashes in the protein, we also performed calculations that considered both the DNA-bound state and the ligand-bound state using a multi-state design algorithm [24]. Finally, we also ran calculations that included an energy bias to favor the wild-type residue. The lowest energy sequences from these four calculations (single-state biased, single-state non-biased, multi-state biased, and multi-state non-biased) were analyzed, and used as a guide to compile a set of ten mutants (Table 2.1) for *in vitro* testing. A more detailed description of the computational methods used can be found in the Materials and Methods section.

### 2.3.2 *In Vitro* Screening of Generated Sequences

We first decided to validate function of the wild-type protein. This was done by placing green fluorescent protein (GFP) downstream of the qacA promoter sequence ($P_{QacA}$). While we observed a hundred-fold decrease in fluorescence in cells containing plasmids encoding the wild-type qacR gene in addition to $P_{QacA}$–GFP, addition of berberine, a native qacR inducer, yielded no observable difference in fluorescence (Figure 2.4). We hypothesized that the inducer was not getting into the cells due to the differences in cell wall permeability between gram-positive and gram-negative

bacteria. Because of this, we decided to use an *in vitro* transcription-translation (TX-TL) system to test the mutants [25].

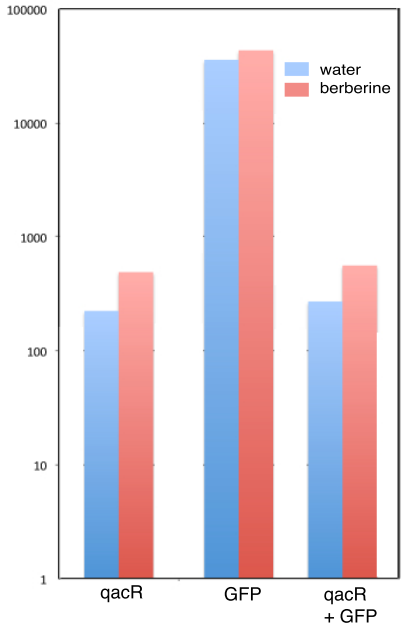
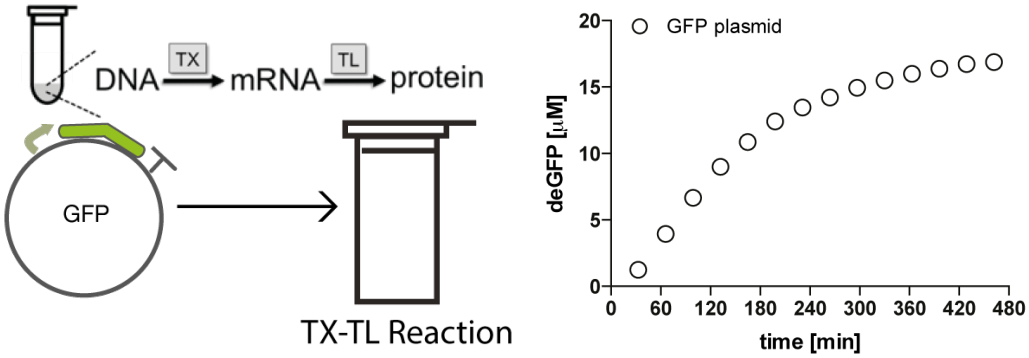



Figure 2.4: **Initial qacR induction test.** Fluorescence of cells encoding wild-type qacR was compared in the presence and absence of berberine, a native qacR inducer. While we observed repression upon the addition of qacR, we did not observe induction when berberine was added.




(a) (b)

Figure 2.5: **TX-TL allows us to prototype circuits *in vitro*** (a) TX-TL contains the transcriptional and translational machinery allowing you to express proteins in the reaction. Adding plasmid DNA encoding proteins allows for their expression and detection in TX-TL. (b) deGFP expression from a TX-TL reaction with plasmid encoding GFP. *(b) adapted from [26]*

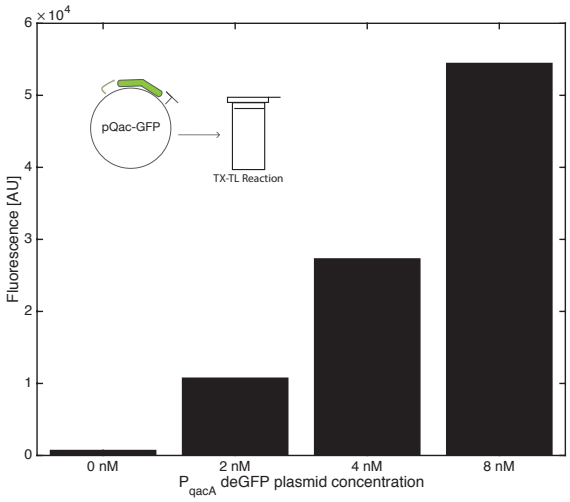
The TX-TL system contains whole cell lysate from BL21 *E. coli* Rosetta 2, with no endogenous mRNA or DNA. A TX-TL reaction is typically done in a 10$\mu$L reaction volume and contains the cell extract, an energy solution consisting of amino acids, nucleotides and 3-PGA, and DNA. It

contains the transcriptional and translational machinery of *E. coli*, allowing one to express proteins by adding plasmid DNA encoding genes one wants expressed (Figure 2.5). Protein concentration can be controlled directly by varying the amount of DNA placed in the reaction. One can execute genetic networks in a TX-TL reaction by adding plasmids that contain proteins that interact. The TX-TL prototyping provides advantages over *in vivo* circuit testing, as it allows us to control protein levels without worrying about promoter and ribosomal binding site strength. Cell wall permeability and protein toxicity are also not issues in the TX-TL system.

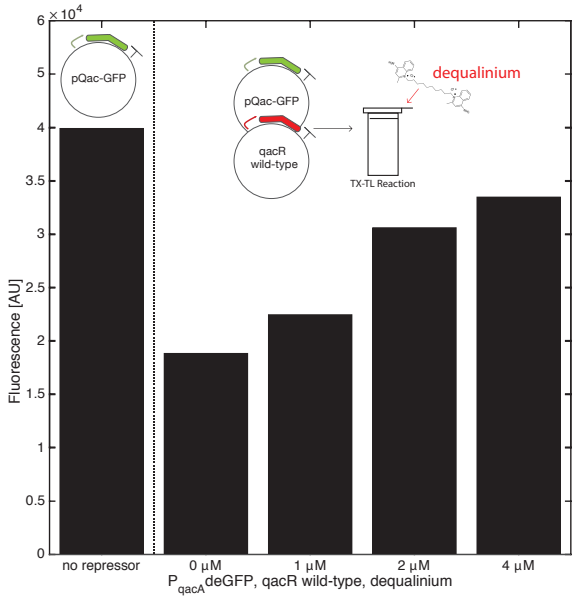
We first tested the wild-type protein in our TX-TL system. We observed an increase in GFP fluorescence as we increased the concentration of plasmid encoding $P_{QacA}$–GFP from 2 nM to 8 nM (Figure 2.6A). The addition of plasmid encoding the qacR repressor to the system resulted in a decrease in fluorescence. Because of the high autofluorescence of berberine, we used dequalinium, a colorless native qacR inducer. The addition of dequalinium resulted in an increase in fluorescence until about 85% of the fluorescence when no DNA encoding repressor was present (Figure 2.6B). These results demonstrated a functional wild-type qacR repressor in TX-TL. After validating the function of wild-type protein in TX-TL, we used the system to look at the functionality of the qacR mutants.

None of the initial mutants showed any repression of GFP fluorescence. We analyzed the ligand-bound and DNA-bound computational models of one of the qacR mutants that contained only three amino acid substitutions from a qacR mutant that was previously shown to be functional by Peters *et al.* [21]. The computational model showed the potential for some mutations to cause steric clashes in the DNA bound state (Figure 2.7). We created a second library reverting either the 50th and 54th positions (A50F/W54L) or the 119th position (Y119L) to their wild-type identity (Table 2.2)

In order to determine if any of the mutants of our library warranted further characterization, we performed a rapid screen of 17 qacR mutants in TX-TL (Figure 2.8). Plasmids containing DNA that encoded each of the qacR variants or the wild-type qacR sequence were placed into a TX-TL reaction containing either water, dequalinium, or vanillin. QacR activity was monitored by a plasmid that encoded GFP downstream of $P_{QacA}$. Two of the mutants, qacR2 and qacR5, displayed an increase

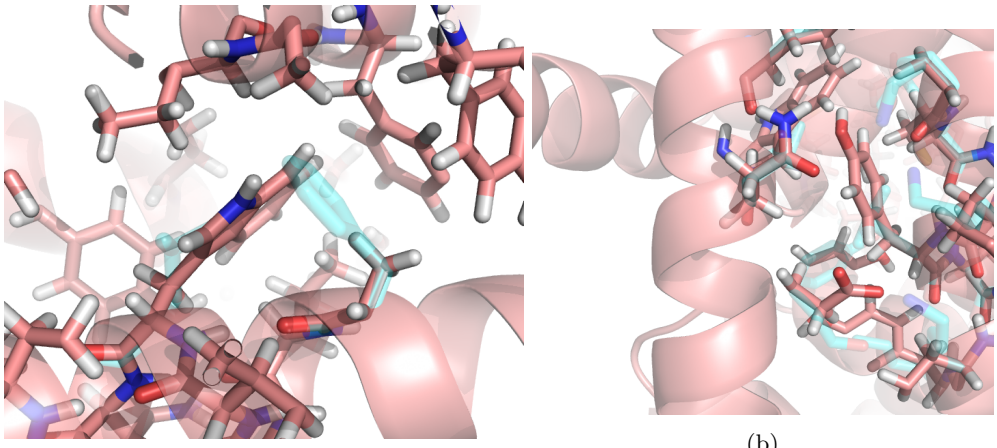



(a)




(b)

Figure 2.6: **Validation of TX-TL screening.** (a) GFP signal after three hours of a TX-TL reaction. Plasmid encoding GFP downstream of the native qac promoter was added to the TX-TL platform. Higher concentrations of plasmid yielded more GFP signal. (b) Response of wild-type qacR to dequalinium. DNA encoding GFP and wild-type qacR was added to the TX-TL system. Increasing fluorescent signal is observed with increasing concentrations of dequalinium. The highest fluorescent signal is observed when there is no repressor in the system, demonstrating the ability of TX-TL to test for qacR repression and de-repression.



(a) (b)

Figure 2.7: **Steric clashes from computational model.** We looked at a computational model (red) of the DNA-bound qacR structure that was three mutations away from a qacR mutant that was previously shown to be functional. The wild-type residues from the structure are shown in cyan. These mutations could have been causing steric clashes in the DNA-bound state. We observed a potential steric clash between the tryptophan (position 54) and tyrosine (position 119) mutations. Based on this model we grouped two of the mutations together and reverted them to their wild-type identity F50A/L54W (a) and Y119L (b), to relieve this potential clash.

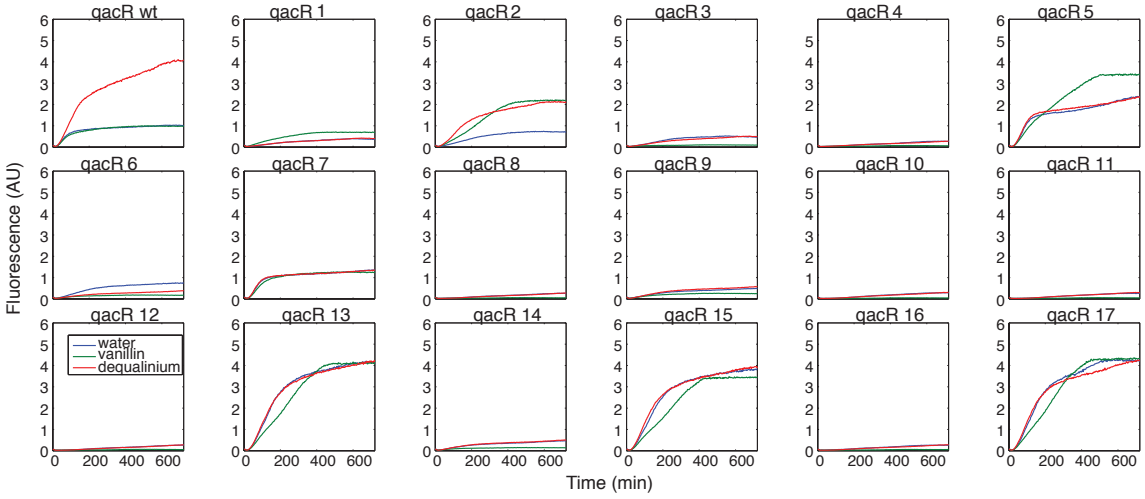



Figure 2.8: **Timetrace of qacR mutant screen.** GFP fluorescence time traces for TX-TL reactions set-up to screen for qacR mutants that were sensitive to vanillin. Reactions contained DNA encoding a qacR variant, T7 RNA polymerase, a fluorescent reporter, and either water, dequalinium, or vanillin.

in fluorescence in the presence of vanillin and dequalinium over water (Figure 2.9). QacR1 was not selected due to its low signal without inducer. We focused on these two mutants for further *in vitro* and *in vivo* characterization.

Table 2.2: **QacR mutants tested** List of the amino acid mutations by position of the second set of qacR mutants tested in the second TX-TL screen. The amino acid identities of the wild-type protein for the positions considered are also shown.

| **Protein** | **Amino Acid Position** | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 50 | 54 | 57 | 58 | 61 | 86 | 89 | 90 | 93 | 96 | 99 | 102 | 116 | 119 | 120 | 126 | 154 | 157 | 161 |
| **qacR-wt** | F | L | E | E | W | S | T | E | Y | Q | I | F | M | L | E | A | N | N | T |
| qacR1 | F | L | Q | L | Y | S | T | Q | Y | M | Q | S | Q | Y | Q | A | M | Q | M |
| qacR2 | F | L | Q | L | Y | S | T | Q | Y | Q | Q | F | Q | Y | Q | A | M | L | M |
| qacR3 | F | L | Q | L | Y | S | T | Q | Y | Q | I | S | Q | Y | Q | A | M | L | M |
| qacR4 | F | L | Q | L | Y | S | T | Q | Y | M | Q | Q | M | Y | Q | A | M | Q | M |
| qacR5 | F | L | E | E | Y | S | T | Q | Y | M | Q | S | Q | Y | Q | A | N | N | T |
| qacR6 | F | L | E | E | Y | S | T | Q | Y | Q | Q | F | Q | Y | Q | A | N | N | T |
| qacR7 | F | L | E | E | Y | S | T | Q | Y | Q | I | S | Q | Y | Q | A | N | N | T |
| qacR8 | F | L | Q | L | W | S | T | Q | Y | M | Q | S | Q | Y | Q | A | M | Q | M |
| qacR9 | A | W | Q | L | Y | S | T | Q | Y | M | Q | S | Q | L | Q | A | M | Q | M |
| qacR10 | A | W | Q | L | Y | S | T | Q | Y | Q | Q | F | Q | L | Q | A | M | L | M |
| qacR11 | A | W | Q | L | Y | S | T | Q | Y | Q | I | S | Q | L | Q | A | M | L | M |
| qacR12 | A | W | Q | L | Y | S | T | Q | Y | M | Q | Q | M | L | Q | A | M | Q | M |
| qacR13 | A | W | E | E | Y | S | T | Q | Y | M | Q | S | Q | L | Q | A | N | N | T |
| qacR14 | A | W | E | E | Y | S | T | Q | Y | Q | Q | F | Q | L | Q | A | N | N | T |
| qacR15 | A | W | E | E | Y | S | T | Q | Y | Q | I | S | Q | L | Q | A | N | N | T |
| qacR16 | A | W | Q | L | W | S | T | Q | Y | M | Q | S | Q | L | Q | A | M | Q | M |
| qacR17 | A | W | E | E | W | S | T | Q | Y | Q | I | F | M | L | Q | A | N | N | T |

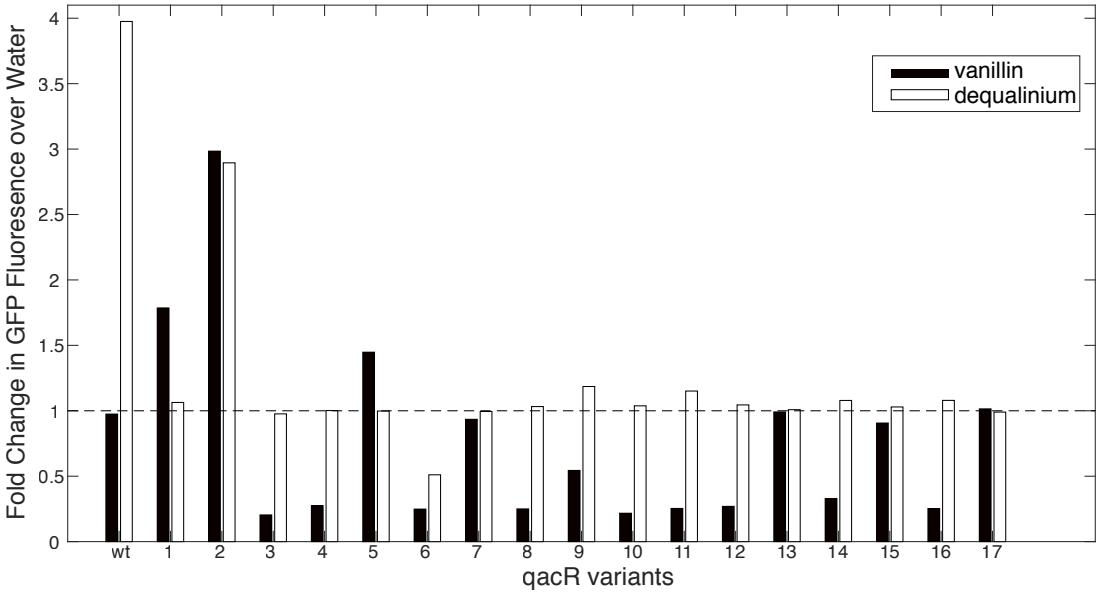



Figure 2.9: ***In vitro* TX-TL screen of qacR mutants found potential candidates for further testing.** Fold change in maximum fluorescence between water and inducer for qacR mutants. Seventeen qacR mutants were screened using TX-TL. Plasmids containing DNA encoding each of the qacR variants were placed into the system along with water, dequalinium (native qacR inducer), and vanillin. To monitor qacR response, a plasmid encoding GFP downstream of the native qacA promoter was also added to the system. qacR2 and qacR5 were selected for further characterization. qacR1 was not selected due to low signal (Figure 2.8)

### 2.3.3 Further *In Vitro* Testing of QacR2 and QacR5

In order to verify the response of qacR2 and qacR5 to vanillin, we performed more extensive TX-TL tests on the mutants. TX-TL reactions were set up with a constant amount of reporter ($P_{QacA}$–deGFP) plasmid and either no repressor (water), or plasmids encoding wild-type qacR, qacR2, or qacR5. Reactions were incubated for 85 minutes at 29 degrees Celsius to produce the repressor protein. This bulk reaction was then added to solution containing dequalinium, vanillin, or water. We monitored the rate of GFP production between the first and third hours of the reaction, where the rate of protein production appeared linear.

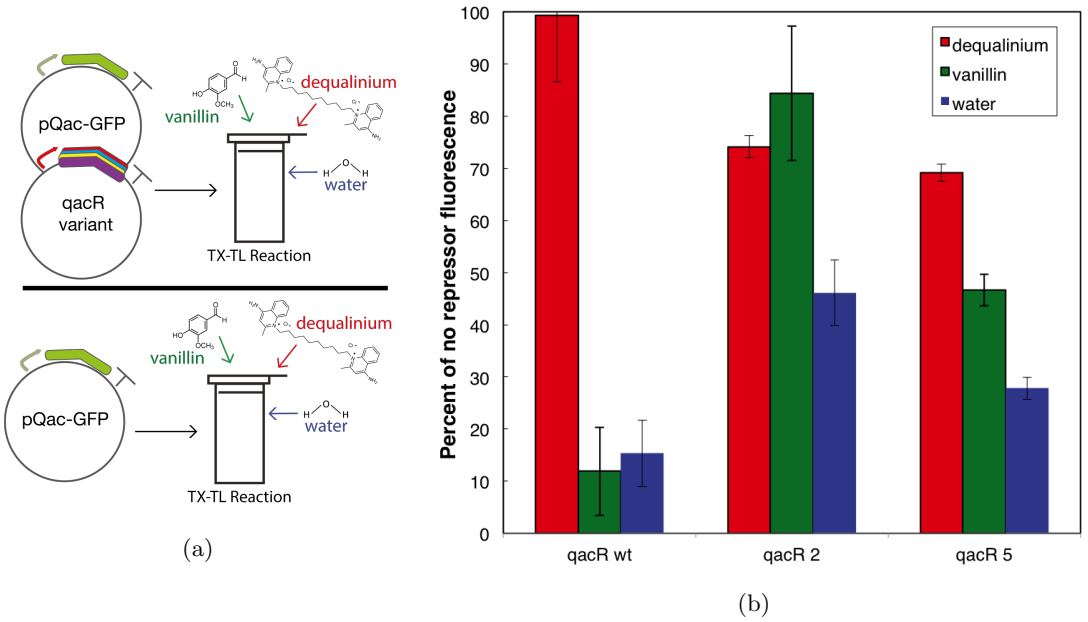



Figure 2.10: ***In vitro* testing of qacR2 and qacR5.** (a) TX-TL reactions were set up with plasmids containing GFP downstream of a qacR sensitive promoter in the presence and absence of plasmids containing the qacR variants under different inducer conditions. To account for inducer toxicity to the TX-TL reaction, and resource limitations from the production of the qacR repressor, we normalized each condition to the reactions that only contained the GFP plasmid under different inducer conditions. (b) Ratio of the rate of GFP production between TX-TL reactions with and without repressor DNA. 10 $\mu$M of dequalinium and 5 mM of vanillin was used to induce the production of GFP for each of the qacR variants tested.

Figure 2.10 shows the ratio of GFP fluorescence between the case where there is no repressor, and each of the repressors tested with the different inducers. The wild-type qacR is able to inhibit the production of fluorescence to around 15% of its maximum value. The mutants are less efficient at repressing the production of GFP. Three times and four times more repressor DNA was added

to the reactions of qacR2 and qacR5, respectively. In spite of the additional DNA, we do not observe the same level of repression that we see with the wild-type protein. Wild-type qacR is well induced by the native inducer, and we observed full derepression at the dequalinium concentration used. Induction of qacR2 and qacR5 with dequalinium is also observed, although to a lesser degree than the wild-type protein. QacR2 and qacR5 display a response to vanillin at the concentration we tested, while no response to vanillin was detected for the wild-type protein. The mutations introduced to the protein decrease the ability of the mutants to repress DNA. This could be due to protein instability, or due to a weaker protein-DNA interaction. However, these mutations also increase the sensitivity of the mutants to vanillin, allowing their response to be detectable in our *in vitro* platform.

We assumed that the maximum amount of GFP fluorescence that can be achieved for a specific inducer condition was when there is no repressor present. This takes into account potential toxicity of the inducer to the TX-TL reaction. The factors that can affect the ability of the particular repressor to reach this the no repressor case are resource limitations due to additional load from the production of the repressor DNA, and response of the repressor to the inducer in the reaction. We expect that resource limitations would have a negative effect on the ability of the repressor to reach the maximum fluorescence level. Conversely, response to repressor should have a positive effect in reaching the maximum fluorescence level.

### 2.3.4 *In Vivo* Testing of QacR2 and QacR5

In order to further characterize the qacR mutants, and to see if we could detect vanillin in a more complex system, we decided to test the *in vivo* response of the qacR variants to vanillin. Plasmids containing genes that encode the wild-type qacR sequence, qacR2 or qacR5 downstream of $P_{Tet}$ and GFP downstream of $P_{QacA}$, were cloned into DH5$\alpha$Z1 cells (Figure 2.11). We performed some testing of the response of the wild-type circuit to dequalinium *in vivo*. In our limited testing, we did not detect a significant difference in response in the wild-type circuit when induced with vanillin or dequalinium at the dequalinium concentrations we tested. For each of the qacR variants,

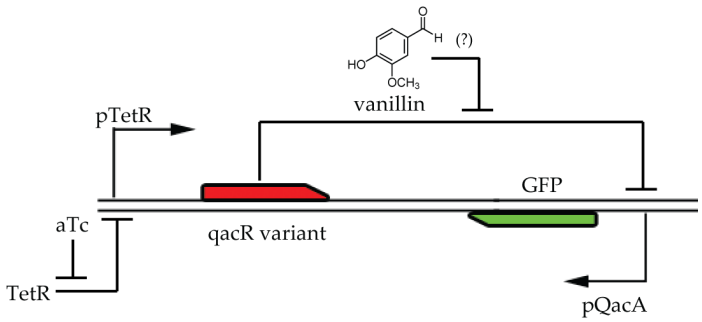
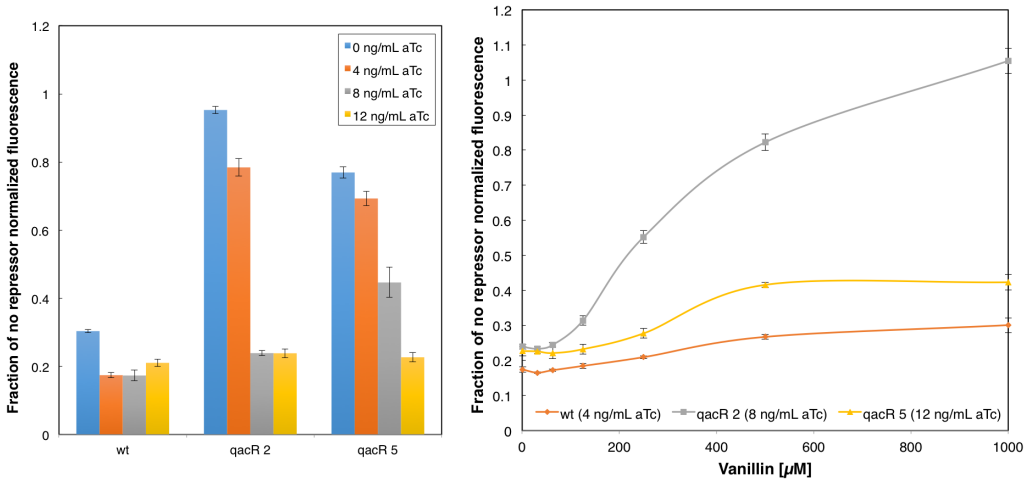



Figure 2.11: **Circuit layout for *in vivo* tests.** Genes encoding GFP under the control of the native qac promoter, and our QacR designs under the control of a tet-inducible promoter were placed in a single plasmid and transformed into DH5$\alpha$Z1 cells. qacR levels were controlled using aTc for varying vanillin concentrations. Candidate designs that are responsive to vanillin should show an increase in fluorescence with increasing vanillin concentrations.

we compared differences in fluorescence signal across increasing vanillin concentrations. We tested different repressor concentrations by varying the amount of anhydrous tetracycline (aTc) in the system. Similar to the *in vitro* experiments, and in order to get an idea for the maximum fluorescence the system could achieve, we grew cells that only contained GFP downstream of $P_{QacA}$ without any repressor. Cells that were grown in higher aTc concentrations had a lower measured optical density (OD), indicating a slower doubling time. We hypothesize that this is due to the toxicity of the qacR repressor to the *E. coli* strain. Since qacR is not a native protein, it is possible that qacR is binding to locations in the *E. coli* genome. Interestingly, the differences in optical density measurements become less pronounced with increasing vanillin concentration, suggesting that vanillin may provide a mitigating effect to this toxicity. In order to account for differences in OD, fluorescence measurements were normalized to OD.

The lowest OD measurements were observed for cells encoding the wild-type qacR at 12 ng/mL aTc, where very little growth was observed for cells expressing the wild-type protein. At this aTc concentration, all of the cells expressing repressor exhibited lower optical densities when compared to cells that were only expressing fluorescent protein. The differences in optical density are less pronounced at lower aTc concentrations. When no aTc is present in the system, cells at the higher vanillin concentrations had lower ODs. At higher aTc concentrations, cells at higher vanillin concentrations had higher ODs. This implies that both the vanillin concentration and the expression

of the repressor have an effect on cellular growth. The optical densities for the cells at different aTc and vanillin concentrations are shown in Tables 2.8-2.11.




(a) (b)

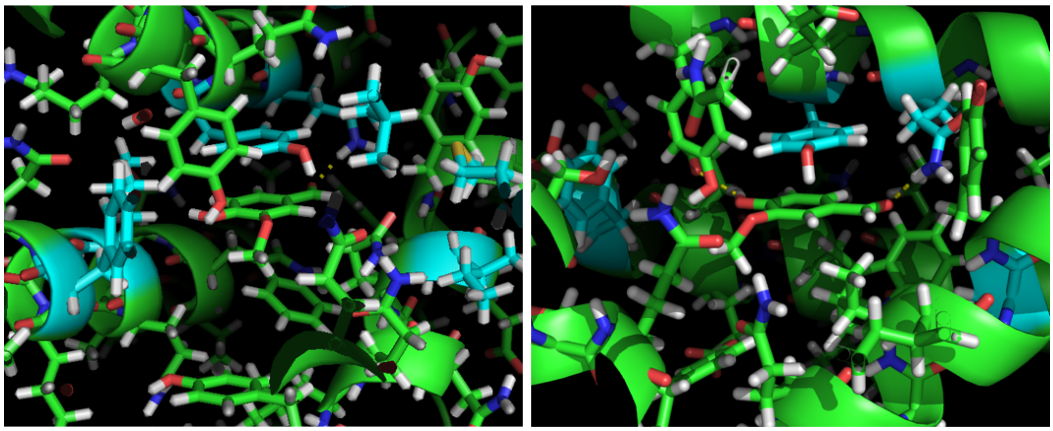
Figure 2.12: ***In vivo* response of qacR to vanillin.** Cells expressing GFP without any repressor were used as a control to normalize for differences in fluorescence due to aTc and vanillin levels. (a) All of the proteins are able to repress the expression of GFP. The wild-type protein is able to inhibit the expression of GFP at lower aTc concentrations, while higher aTc concentrations are necessary for the mutants to achieve a similar level of repression. (b) QacR mutants respond to vanillin in a concentration dependent manner.

Figure 2.12A shows the effect of increasing the aTc concentration on the fluorescence of cells in the absence of vanillin. Similar to the *in vitro* tests, fluorescence was normalized to the no repressor case. Increasing the aTc concentration decreased the fluorescence of cells in the absence of vanillin, confirming that the qacR mutants are able to repress the expression of GFP at higher protein concentrations.

The response of wild-type qacR, qacR2, and qacR5 to increasing vanillin concentrations is shown in Figure 2.12B. The response curves for each protein are plotted for the minimum aTc concentration such that maximum GFP repression is observed. This corresponds to aTc concentrations of 4, 8, and 12 ng/mL for wild-type qacR, qacR2, and qacR5 respectively. This is consistent with the *in vitro* data that more qacR2 and qacR5 DNA was required to repress the expression of GFP. Similar to the *in vitro* tests, we expect the ability of the cell to reach the maximum fluorescence level to be dependent on its response to inducer, and toxicity from vanillin and qacR. Indeed, cells expressing

the qacR mutants exhibited an increase in fluorescence with increasing vanillin levels demonstrating that they are capable of sensing vanillin. While all three proteins appear to be sensitive to vanillin, the mutants exhibit a marked increase in sensitivity to vanillin. QacR2 displays a response that goes from approximately 20 % of the fluorescence of the cells not expressing any repressor to matching the fluorescence of the non-repressed cells at 1 mM vanillin. QacR5 saturates at around 40 % of the fluorescence of the non-repressed cells. This correlates with the *in vitro* data that show qacR2 achieving close to the non-repressed fluorescence, with qacR5 less sensitive to vanillin (Figure 2.10). Figure 2.16 shows the vanillin dosage response of qacR wild-type, qacR2, and qacR5 for different aTc concentrations tested.

### 2.3.5 Analysis of Mutations of QacR2 and QacR5

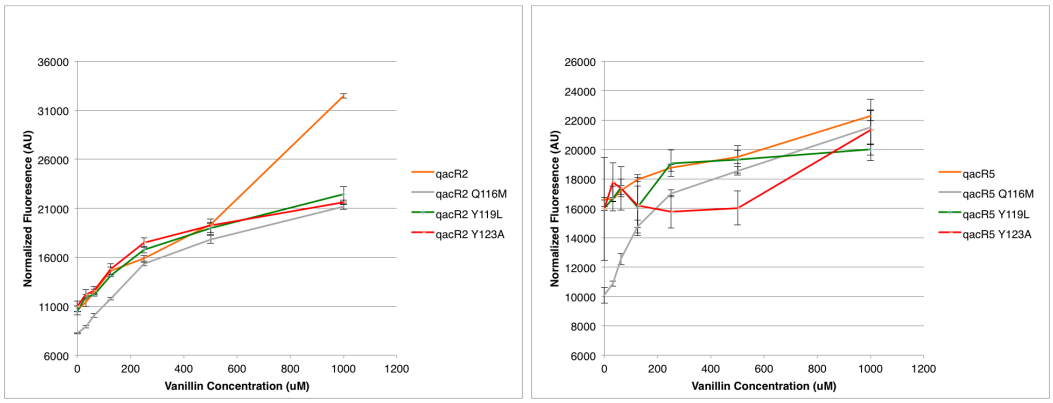


(a) (b)

Figure 2.13: **Model of vanillin bound to qacR2 and qacR5.** Model from computational design software of vanillin bound to (a) qacR2 and (b) qacR5 centered on vanillin. Residues that are mutated from wild-type are colored in cyan. In both qacR2 and qacR5 key mutations seem to be a tyrosine (position 119) above the vanillin forming a pi-stacking interaction, and a glutamine (position 116) that forms a potential hydrogen bond with the aldehyde group of the vanillin (potential interaction denoted by the yellow dotted line).

In order to evaluate the contribution of the mutations from the computational design on the protein's increased sensitivity to vanillin, we analyzed the computational models of the vanillin bound to qacR2 and qacR5. In both cases, there are three key interactions that the CPD algorithm suggests: (1) a hydrogen bond between the aldehyde group of the vanillin and Q116 (mutated from

methionine), (2) a pi-stacking interaction between vanillin and Y119 (mutated from leucine), and (3) hydrogen bond interactions between the hydroxyl group of Y123 and the ether and hydroxyl groups of vanillin (wild-type residue) (Figure 2.13). We examined the role of the first two interactions on vanillin sensitivity by measuring the response of point mutants that reverted residues 116 and 119 back to their wild-type identity. To investigate the role of the third interaction, we examined Y123A point mutants of qacR2 and qacR5. Figure 2.14 shows the normalized fluorescence for increasing concentrations of vanillin. None of the point mutations seemed to significantly affect vanillin sensitivity of either qacR2 or qacR5. This suggests that either the combination of the remaining interactions is still able to confer an increase in vanillin sensitivity for the protein, or that the mechanism of vanillin sensitivity is different from what the computational model suggests.




(a) (b)

Figure 2.14: **Vanillin dosage response curves for point mutants of qacR2 (a) and qacR5 (b).** GFP normalized to optical density for varying vanillin concentrations.

### 2.3.6 Framework Enables Engineering of Sensors through Rational Reduction of Design Space

The framework developed— a combination of sequence generation using computationally-aided design, preliminary screening with TX-TL, and *in vitro* and *in vivo* validation—can be used for other small molecule targets, potentially facilitating the design of more sensors in synthetic circuits. While it is possible that the computational model of vanillin binding was inaccurate, the computational design provided value in drastically reducing the amount of sequences to test into a number that

was experimentally tractable. Without the computational design to reduce the size of the design space, we would not have had starting points to attempt the engineering of a vanillin sensor.

The use of the *in vitro* cell-free system in a preliminary screen provides many advantages. It allows the screening of more mutants in a shorter amount of time. The simpler system also reduces the number of variables to consider. Factors such as cell membrane permeability and cell growth do not need to be considered during this part of the screen. Repressors whose native inducers cannot enter the target organism can be used as starting points with the cell-free system. Finally, we can use this framework to target molecules that are known to be toxic to cells.

### 2.3.7 Future Directions

As a result of this process, we now have functional vanillin sensors that can be used in a feedback circuit that dynamically responds to vanillin. QacR2 and qacR5 can be used as a starting point for a synthetic circuit that responds to vanillin concentrations. While we only tested the protein in *E. coli*, recent work has developed a process that facilitates the transfer of prokaryotic transcription factors into eukaryotic cells, increasing the flexibility of the molecules for use in metabolic engineering [27]. The first step would be to see if you can use a similar process to obtain vanillin responsive promoters in *S. cerevisiae*.

If the designed promoters are able to show a dynamic response to vanillin concentrations, we can use this system in a feedback loop that will allow *S. cerevisiae* to dynamically respond to vanillin. We aim to do this by placing these promoters downstream of genes that we have reason to believe can increase vanillin tolerance in yeast, specifically PDR12, YHK8, and ADH6. These genes were discovered as part of a screen that looked at the increased sensitivity of a yeast knockout library to phenolic byproducts of lignin degradation [11]. Some of these genes are predicted to be efflux pumps (YHK8 and PDR12) and thus metabolically expensive to produce [28]. Before using these in a feedback circuit, we would first have to verify that these genes provide a benefit to the growth of the yeast in the presence of vanillin. We would then aim to show that it there is a growth advantage to the vanillin mediated expression of these genes as opposed to their constitutive expression.

Challenges in this part of the project include balancing gene expression levels to make sure that the components are in the correct operating regimes for the vanillin concentration encountered. Furthermore, it remains to be seen whether the sensors developed will have the required dynamic range or sensitivity for a functional feedback circuit; however, if a better sensor is needed then these proteins can be used as a starting point for directed evolution in order to obtain a sensor with the desired properties.

## 2.4 Materials and Methods

### 2.4.1 Computationally Aided Selection of Mutant Sequences

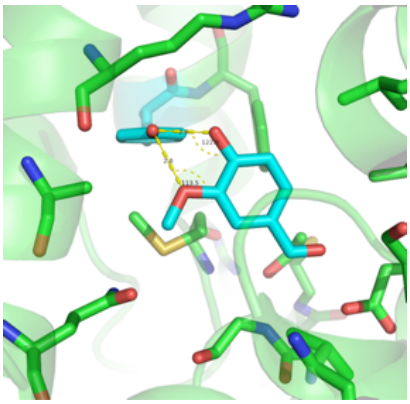


Figure 2.15: **Crystal Structure of Vanillin Bound to a Protein (PDBID 2VSU)**. A model of vanillin binding was constructed by building vanillin off a native tyrosine residue similar to what we observe here. The hydroxyl group of the tyrosine has potential hydrogen-bonding interactions with the methoxy and hydroxyl groups of the vanillin.

An *in silico* model of vanillin was constructed using the Schrödinger software suite. Partial charges for vanillin were computed using Optimization in Jaguar version 7.6 [29] using HF/6-311G** as the basis set. Vanillin rotamers were chosen by looking at the ideal angles for the carbon hybrid orbitals. A model of an idealized vanillin binding pocket was designed by looking at the protein data bank for proteins that bound small molecules similar to vanillin, specifically PDBID 2VSU. Models of vanillin in the qacR binding pocket were generated using the Phoenix Match algorithm [23].

Vanillin was built off a native tyrosine residue (Y123), the primary interaction considered for the algorithm was a hydrogen bonding interaction between the hydroxyl group of the tyrosine with

the methoxy and hydroxyl groups of the vanillin (Figure 2.15). We modified the energy function to include an energy bias of −100 kcal/mol for potential pi-stacking interactions between vanillin and tyrosine, phenylalanine, or tryptophan residues. We also included an energy bias hydrogen-bonding interactions with the methoxy, hydroxyl, and aldehyde groups of vanillin with serine, threonine, tyrosine, glutamine, or asparagine residues. Tables 2.3-2.6 show the geometry ranges used for the energy bias function. The Phoenix Match algorithm was asked to return potential vanillin binding locations that contained interaction with the native tyrosine, at least one pi-stacking interaction, and at least two other hydrogen bonding interactions. Solutions from the algorithm were grouped together and resulted in four potential spots for vanillin. These locations were used as vanillin "rotamers" for computational protein design.

Table 2.3: **Geometry definitions for tyrosine to vanillin methoxy and hydroxyl group interaction.** Geometry definitions for the bias function for the interaction between vanillin and the wild-type tyrosine (position 123) in qacR. [A, B] indicates the centroid of the coordinates of the two atoms specified

| | Vanillin Atoms | Residue Atoms (Tyrosine) | Range |
|---|---|---|---|
| Distance | [O1,O3] | CZ | 2.8–3.0 Å |
| Angle 1 | O3, [O1,O3] | CZ | 69°–110° |
| Angle 2 | [O1,O3] | CZ, CE1 | 100°–140° |
| Dihedral 1 | C6, O3, [O1,O3] | CZ | 120°–240° |
| Dihedral 2 | [O1,O3] | CZ, CE1, CE2 | 160°–200° |
| Dihedral 3 | O3, [O1,O3] | CZ, CE1 | 0°–360° |

Table 2.4: **Geometry definitions for pi-stacking interaction above vanillin.** Geometry definitions for the bias function for the interaction between vanillin and an aromatic residue in qacR. [A, B] indicates the centroid of the coordinates of the two atoms specified. Atom identities for specific amino acids: tyrosine/phenylalanine: A–CE1, B–CD2, C–CG, tryptophan: A–CE2, B–CD2, C–CE3

| | Residue Atoms | Vanillin Atoms | Range |
|---|---|---|---|
| Distance | [A,B] | [C1,C4] | 3.0–4.0 Å |
| Torsion | [A,B] | [C1,C4], C6, C1 | 0°–180° |
| Plane | [A,B,C] | C1, C3, C5 | 0°–40° |

Table 2.5: **Geometry definitions for pi-stacking interaction below vanillin.** Geometry definitions for the bias function for the interaction between vanillin and an aromatic residue in qacR. [A, B] indicates the centroid of the coordinates of the two atoms specified. Atom identities for specific amino acids: tyrosine/phenylalanine: A–CE1, B–CD2, C–CG, tryptophan: A–CE2, B–CD2, C–CE3

| | Residue Atoms | Vanillin Atoms | Range |
|---|---|---|---|
| Distance | [A,B] | [C1,C4] | 3.0–4.0 Å |
| Torsion | [A,B] | [C1,C4], C6, C1 | 180°–360° |
| Plane | [A,B,C] | C1, C3, C5 | 0°–40° |

Table 2.6: **Geometry definitions for hydrogen bonds between qacR and vanillin aldehyde.** Geometry definitions for the bias function for the interaction between vanillin and a hydroxyl containing residue in qacR. Atom identities for specific amino acids: tyrosine: A–OH, B–HH, serine: A–OG, B–HG, threonine: A–OG1, B–HG1, glutamine: A–NE2, B–1HE2, or 2HE2, asparagine: A–ND2, B–1HD1, or 1HD2

| | Residue Atoms | Vanillin Atoms | Range |
|---|---|---|---|
| Distance | B | O2 | 1.6–2.0 Å |
| Angle 1 | A, B | O2 | 155°–180° |
| Angle 2 | B | O2, C7 | 100°–140° |

Monte Carlo with simulated annealing [30] and FASTER [31] were used to sample conformational space. A backbone independent conformer library with a 1.0 Å resolution was used for the designed residues [23]. Designed residues were chosen by compiling a list of amino acid residues within 15 Å of vanillin. Table 2.7 shows the amino acid design positions, and the allowed amino acid residues for each position. Allowed amino residues for each site were selected by visually inspecting the qacR cyrstal structure with the potential vanillin binding locations. Rotamer optimization was allowed for other residues in the 15 Å shell in which mutations were not allowed. Computational models of qacR with vanillin present were scored using the PHOENIX forcefield with the inclusion of an additional geometry bias term that favored pi-stacking and hydrogen bonding interactions [23] that we used to find potential vanillin active sites. We considered solutions that both included and excluded $-20$ kcal/mol wild-type bias term in the energy function.

### 2.4.2 Cell-Free *In Vitro* Transcription-Translation System and Reactions

The transcription-translation reaction consists of crude cytoplasmic extract from BL21 Rosetta 2 *E. coli* [25]. Preliminary tests were done with plasmids and inducers at the specified concentrations. For the initial screen, the qacR mutants were downstream of a T7 promoter. TX-TL reactions were run with 2 nM of the plasmid encoding the qacR variant, 0.1 nM plasmid encoding T7 RNA polymerase, and 8 nM plasmid encoding $P_{QacA}$–deGFP. Vanillin was added at a concentration of 2.5 mM and dequalinium was added at 10 $\mu$M.

For the *in vitro* tests to further characterize the hits, plasmids encoding qacR2 or qacR5 downstream of a tet-responsive promoter were used along with a plasmid encoding deGFP downstream

Table 2.7: **Design Positions for CPD.** Residues around the potential vanillin binding sites were examined in the context of the crystal structure and allowed amino acids were chosen based on the location of the specific residue and the potential interactions with vanillin.

| Residue ID | Wild-Type Identity | Allowed Amino Acids |
|---|---|---|
| 50 | F | FWYILMA |
| 54 | L | STNQYWF |
| 57 | E | STNQYWF |
| 58 | E | WFYMLIQNST |
| 61 | W | FWY |
| 86 | S | FWYILM |
| 89 | T | FWYILMQNSVT |
| 90 | E | FWYILMQNSTV |
| 96 | Q | WFYMLVIQNST |
| 99 | I | FWYSTNQ |
| 102 | F | WYQNST |
| 116 | M | FWYIQNST |
| 119 | L | FWYQNST |
| 120 | Q | FWYQN |
| 126 | A | FWYILMV |
| 154 | N | WFYMLVIQNST |
| 157 | N | QVILFWY |
| 162 | T | WFYMLVIQNST |

of a qac-responsive promoter. Plasmids were prepared using the Macherey-Nagel NucleoBond Xtra Midi/Maxi Kit. Plasmid DNA was eluted in water and concentrated by vacufuge to the desired concentration. TX-TL reactions were set up as follows: 5 $\mu$L of buffer, 2.5 $\mu$L of cell extract and 1.5 $\mu$L repressor DNA at a specific concentration was mixed and incubated at 29 C for 75 minutes to facilitate the production of repressor DNA. This mix was then added to a mixture of 1 $\mu$L deGFP plasmid and 1 $\mu$L of an inducer stock. Measurements were made in a Biotek plate reader at 3 minute intervals using excitation/emission wavelengths set at 485/525 nm. Stock repressor plasmid concentrations were 243 nM , 729 nM , and 972 nM for qacR wild-type, qacR2, and qacR5, respectively. The deGFP plasmid concentration was approximately 397 nM. Inducer concentrations were 5 mM for vanillin, and 10 $\mu$M for dequalinium.

Experimental conditions were done in triplicate and the error bars are the error propagated from the standard deviation of the means.

### 2.4.3 Cell Strain and Media

The circuit was implemented in the *E.coli* cell strain DH5$\alpha$Z1, a variant of DH5$\alpha$ that contains a chromosomal integration of the Z1 cassette [32]. The Z1 cassette constitutively expresses the TetR and LacI proteins. All cell culture was done in optically clear M9ca minimal media (Teknova M8010).

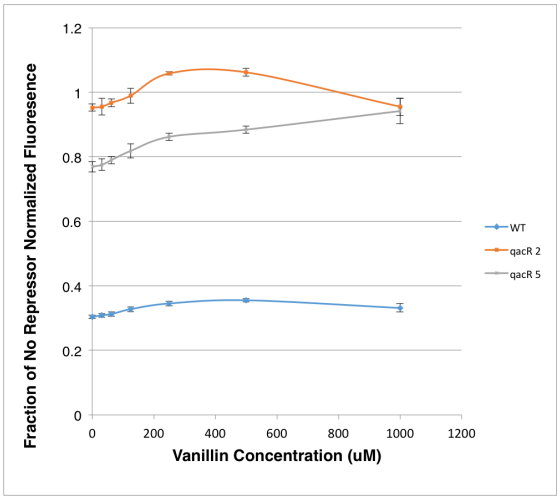
### 2.4.4 Genes and Plasmids

DNA encoding the qacR genes was constructed using overlap extension PCR. Plasmids used contained chloramphenicol resistance with a p15a origin of replication.

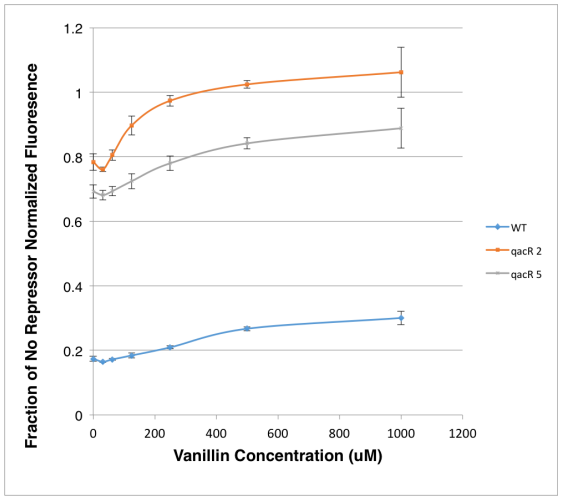
### 2.4.5 *In Vivo* Experiments

Cells were grown in at least two consecutive overnight cultures in M9ca minimal media. On the day of the experiment, overnight cultures were diluted 1:100 and grown for 5 hours to ensure that the cells were in log phase. Cells were then diluted 1:100 into fresh media at the specified experimental condition. Cells were grown in these conditions at 37C for 12–15 hours in Axygen 96 well plates while shaking at 1100 rpm. Endpoint fluorescence was measured by transferring the cells to clear bottomed 96-well microplates (PerkinElmer, ViewPlate, 6005182) . GFP was read at 488/525 with gain 100.

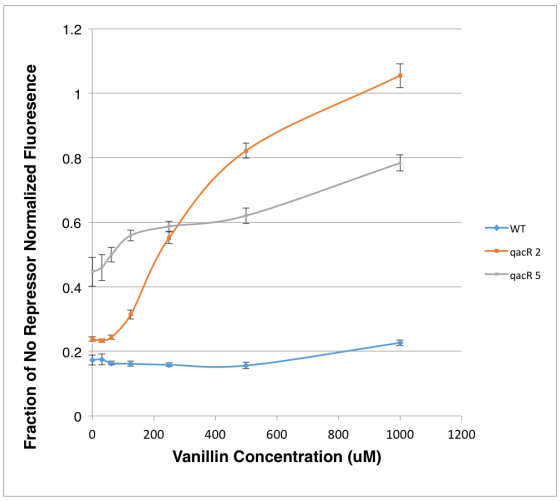
Analysis of the data was done by taking fluorescence readings for each independent well. Experimental conditions of the qacR proteins were done in triplicate and repeats were averaged. Error bars shown are the error propagated originating from the standard deviation of the mean.



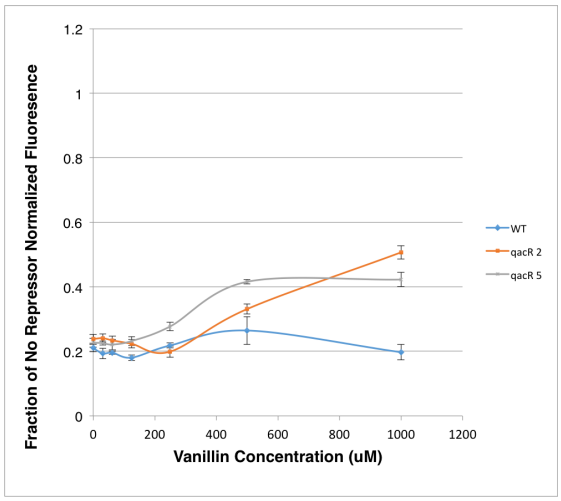

(a) 0 ng/ml aTc.




(b) 4 ng/ml aTc.




(c) 8 ng/ml aTc.




(d) 12 ng/ml aTc.

Figure 2.16: **Vanillin dosage response curves.** Cells expressing GFP without any repressor were used as a control to normalize for differences in fluorescence due to aTc and vanillin levels. Fluorescence was measured for different inducer concentrations. Mutants display an increased sensitivity to vanillin concentrations for most of the range of aTc concentrations tested.

Table 2.8: **OD600 of cells at 0 ng/mL aTc.**

| vanillin (mM) | GFP | | | wild-type | | | qacR2-1 | | | qacR5-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1000 | 0.456 | 0.44 | 0.42 | 0.75 | 0.72 | 0.71 | 0.8 | 0.82 | 0.84 | 0.6 | 0.57 | 0.54 |
| 500 | 0.62 | 0.63 | 0.63 | 0.78 | 0.77 | 0.78 | 0.83 | 0.83 | 0.82 | 0.72 | 0.71 | 0.72 |
| 250 | 0.718 | 0.75 | 0.72 | 0.82 | 0.83 | 0.82 | 0.88 | 0.84 | 0.84 | 0.78 | 0.77 | 0.75 |
| 125 | 0.775 | 0.78 | 0.76 | 0.82 | 0.81 | 0.83 | 0.85 | 0.89 | 0.86 | 0.79 | 0.78 | 0.8 |
| 62.5 | 0.783 | 0.81 | 0.8 | 0.84 | 0.86 | 0.85 | 0.88 | 0.91 | 0.89 | 0.84 | 0.82 | 0.83 |
| 31.25 | 0.823 | 0.83 | 0.82 | 0.84 | 0.85 | 0.85 | 0.87 | 0.88 | 0.88 | 0.85 | 0.82 | 0.83 |
| 0 | 0.835 | 0.86 | 0.82 | 0.85 | 0.86 | 0.85 | 0.89 | 0.91 | 0.9 | 0.86 | 0.84 | 0.84 |
| no aTc | 0.866 | 0.85 | 0.85 | 0.88 | 0.9 | 0.88 | 0.92 | 0.92 | 0.91 | 0.9 | 0.83 | 0.85 |

Table 2.9: **OD600 of cells at 4 ng/mL aTc.**

| vanillin (mM) | GFP | | | wild-type | | | qacR2-1 | | | qacR5-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1000 | 0.52 | 0.49 | 0.48 | 0.66 | 0.66 | 0.68 | 0.78 | 0.79 | 0.78 | 0.6 | 0.6 | 0.59 |
| 500 | 0.63 | 0.64 | 0.63 | 0.73 | 0.73 | 0.75 | 0.83 | 0.84 | 0.83 | 0.71 | 0.7 | 0.72 |
| 250 | 0.78 | 0.74 | 0.74 | 0.75 | 0.76 | 0.79 | 0.89 | 0.85 | 0.85 | 0.76 | 0.77 | 0.77 |
| 125 | 0.76 | 0.8 | 0.78 | 0.74 | 0.77 | 0.75 | 0.86 | 0.89 | 0.88 | 0.8 | 0.79 | 0.79 |
| 62.5 | 0.84 | 0.8 | 0.81 | 0.75 | 0.77 | 0.77 | 0.87 | 0.88 | 0.89 | 0.81 | 0.82 | 0.85 |
| 31.25 | 0.79 | 0.82 | 0.81 | 0.75 | 0.74 | 0.77 | 0.89 | 0.9 | 0.89 | 0.81 | 0.8 | 0.79 |
| 0 | 0.84 | 0.86 | 0.85 | 0.78 | 0.8 | 0.77 | 0.93 | 0.91 | 0.88 | 0.85 | 0.84 | 0.85 |
| no aTc | 0.82 | 0.85 | 0.81 | 0.84 | 0.85 | 0.85 | 0.87 | 0.91 | 0.9 | 0.86 | 0.85 | 0.84 |

Table 2.10: **OD600 of cells at 8 ng/mL aTc.**

| vanillin (mM) | GFP | | | wild-type | | | qacR2-1 | | | qacR5-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1000 | 0.45 | 0.45 | 0.41 | 0.26 | 0.26 | 0.27 | 0.7 | 0.72 | 0.72 | 0.35 | 0.37 | 0.37 |
| 500 | 0.67 | 0.64 | 0.63 | 0.27 | 0.24 | 0.23 | 0.7 | 0.7 | 0.71 | 0.46 | 0.44 | 0.46 |
| 250 | 0.72 | 0.72 | 0.71 | 0.11 | 0.12 | 0.11 | 0.55 | 0.52 | 0.52 | 0.49 | 0.51 | 0.53 |
| 125 | 0.8 | 0.77 | 0.8 | 0.09 | 0.09 | 0.09 | 0.23 | 0.23 | 0.24 | 0.41 | 0.43 | 0.47 |
| 62.5 | 0.87 | 0.82 | 0.81 | 0.08 | 0.07 | 0.08 | 0.14 | 0.15 | 0.15 | 0.38 | 0.38 | 0.42 |
| 31.25 | 0.81 | 0.83 | 0.79 | 0.08 | 0.07 | 0.07 | 0.11 | 0.11 | 0.11 | 0.32 | 0.33 | 0.39 |
| 0 | 0.84 | 0.83 | 0.82 | 0.08 | 0.08 | 0.07 | 0.09 | 0.09 | 0.09 | 0.31 | 0.33 | 0.37 |
| no aTc | 0.8 | 0.79 | 0.8 | 0.84 | 0.86 | 0.83 | 0.89 | 0.88 | 0.88 | 0.83 | 0.8 | 0.82 |

Table 2.11: **OD600 of cells at 12 ng/mL aTc.**

| vanillin (mM) | GFP | | | wild-type | | | qacR2-1 | | | qacR5-1 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1000 | 0.48 | 0.45 | 0.49 | 0.05 | 0.06 | 0.05 | 0.27 | 0.26 | 0.26 | 0.15 | 0.15 | 0.14 |
| 500 | 0.69 | 0.67 | 0.67 | 0.06 | 0.07 | 0.05 | 0.23 | 0.23 | 0.24 | 0.17 | 0.17 | 0.19 |
| 250 | 0.76 | 0.76 | 0.75 | 0.04 | 0.05 | 0.05 | 0.08 | 0.08 | 0.08 | 0.11 | 0.12 | 0.12 |
| 125 | 0.82 | 0.79 | 0.79 | 0.04 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.07 | 0.08 | 0.08 |
| 62.5 | 0.81 | 0.82 | 0.8 | 0.04 | 0.04 | 0.04 | 0.04 | 0.05 | 0.05 | 0.06 | 0.06 | 0.06 |
| 31.25 | 0.82 | 0.84 | 0.82 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.05 | 0.06 | 0.06 | 0.06 |
| 0 | 0.84 | 0.87 | 0.84 | 0.04 | 0.04 | 0.04 | 0.05 | 0.04 | 0.05 | 0.06 | 0.05 | 0.05 |
| no aTc | 0.82 | 0.81 | 0.84 | 0.84 | 0.84 | 0.83 | 0.89 | 0.87 | 0.94 | 0.81 | 0.82 | 0.83 |

# Chapter 3

# Design and Implementation of a Biomolecular Circuit for Tracking Protein Concentration

## 3.1 Introduction

As biomolecular circuits in synthetic biology increase in complexity, the incorporation of closed loop feedback controllers into circuit designs will become necessary for robust performance. In order to do this in a biochemical circuit, a mechanism that can track the difference in the amount of an output to a given reference value is required. With this in mind, we designed a sequestration-based negative feedback circuit, in which one protein tracks the level of another protein, which we consider to be our reference.

This chapter discusses the design, modeling, and biochemical implementation of this circuit. In this circuit, we set the level of our reference, protein A. The presence of protein A conditionally activates the synthesis of protein B, our desired output. Protein B is designed to have regions that interact with protein A, creating a negative feedback loop that shuts down the synthesis of protein B when the levels of A and B are equal.

The design of this circuit was partially inspired by a circuit designed by Franco *et al.* [33] where an *in vitro* RNA transcription circuit coupled the levels of two double stranded DNA species through sequestration and transcriptional regulation. A mathematical model consisting of a set of ODEs is derived from mass action laws and Hill function approximations of protein production. Steady-state

analysis of the model is used to predict parameter sensitivity and experimental behavior. We used the model to explore parameter space and look at the tunability of certain aspects of the circuit such as the ratio between proteins A and B.

The work described in this chapter is primarily derived from a conference paper that was presented in the 2012 American Controls Conference [34]. The work was a collaboration between me, Victoria Hsiao, and Weston Whitaker. I conceptualized the idea of a protein tracker and designed the scaffold-based mechanism with Weston. I then worked on construction and sensitivity analysis of the initial model, and wrote the ACC paper. Victoria worked on the model and did the experimental work for this paper. She continued the work on this project by adding a phosphatase to the model and exploring the predicted effects of adding more response regulator and phosphatase to the protein ratios and maximum scaffold occupancy limit. She performed steady-state and dynamic experiments based on the new model, leading to a publication in *ACS Synthetic Biology* [35]. This chapter mainly covers the conceptualization of the circuit and the framework of the model. A more detailed explanation of the experimental results and refinement of the model can be found in the *ACS Synthetic Biology* paper.

## 3.2 Background

Synthetic feedback regulation is a critical hurdle for the design and engineering of complex synthetic circuits. Recent theoretical work by Ang *et al.* [36] proposed and modeled a two-promoter transcription-level system for implementing integral control. Stapleton *et al.* [37] demonstrated a transcription-level negative feedback circuit in mammalian cells, in which an RNA binding protein repressed translation of its own mRNA. Here, we model and demonstrate a protein-level negative feedback circuit.

Previously, Dueber *et al.* showed that synthetic scaffold proteins could be used to co-localize intermediates in the production of a metabolic product [38]. They designed synthetic scaffolds out of single protein binding domains linked together by short repeating peptide sequences. Further work by Whitaker *et al.* showed that prokaryotic two-component systems could be selectively phosphorylated

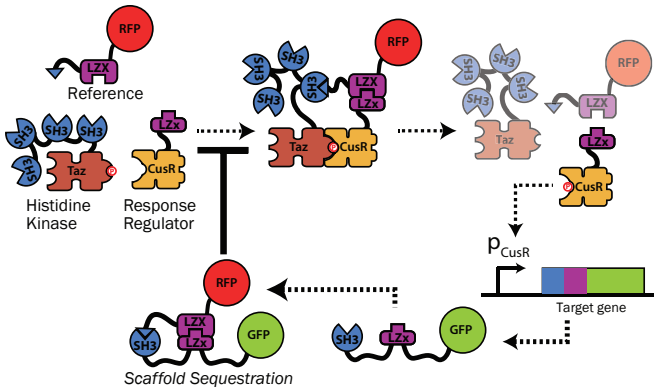



Figure 3.1: A visual schematic of our circuit design. The circuit regulates the production of the amount of target protein with respect to the amount of reference protein. This is done using programmable scaffold domains. Expression of the target is dependent on the amount of free scaffold. The target contains domains that sequester free scaffold, creating a negative feedback loop.

using these eukaryotic scaffolds [39].

Two-component systems are naturally found in bacteria as environmental sensor and response systems [39, 40]. A transmembrane histidine kinase senses environmental conditions, such as osmolarity or pH, and phosphorylates a response regulator protein. The phosphorylated response regulator becomes an active transcription factor, allowing the transcription and translation of target genes.

Building off of this previous work [39], our circuit uses a two-domain synthetic protein scaffold as our reference protein (scaffold) that activates transcription of our target (Fig. 3.1). The target is designed to bind to the scaffold, creating a negative feedback loop via sequestration.

## 3.3 Circuit Description

The circuit consists of four major components: a synthetic scaffold protein (Sc), a histidine kinase (HK), a response regulator (RR), and a target protein. The synthetic scaffold serves as our reference level. It consists of two pairs of small protein-protein binding domains: the SH3-peptide/SH3-domain and the leucine zipper LZx/LZX, which have been previously used and characterized [41]. The HK and RR are commonly found in bacterial signaling systems. In our system, they serve as the mechanism for conditional activation of the target protein based on the reference levels.

In natural systems, there are hundreds of HK/RR pairs that perform signaling functions with minimal crosstalk. Since we desire activation of the RR to be conditional on our reference protein, we have selected a non-cognate pair of histidine kinase and response regulator, known to be orthogonal *in vivo*. Based on previous designs by Whitaker *et al.*, the HK is bound to a SH3 ligand and the RR is bound to a LZx domain, which both bind to Sc [39]. When both HK and RR are bound to the scaffold, the RR becomes an active transcription factor, resulting in the expression of our target protein [39]. The target protein contains domains for both the SH3 ligand and LZx, which are complementary to the scaffold binding domains. This allows the target to out-compete both the HK and RR for binding to the scaffold, effectively sequestering the scaffold, and repressing its own production. We expect that with this circuit design, we can regulate the relative steady-state concentrations of scaffold to anti-scaffold. Figure 3.1 shows a schematic description of our circuit design.

## 3.4 Mathematical Model

We constructed an ODE-based dynamical model of our circuit from the chemical reactions occurring in our system [42]. The model consists of 19 species corresponding to the active and non-active states of the HK and RR, and the different two-species and three-species complexes that can be formed. The model contains 56 different reactions that describe production and degradation of our basic species (Sc, HK, RR, and target), activation of HK and RR, binding of HK and RR to Sc, and sequestration of Sc by our target. With the exception of the formation of our target protein (AS), mass-action kinetics was used to describe the rate laws of the reactions. Parameters were selected from experimental values found in the literature where available [41, 40, 43, 44]. A complete list of chemical reactions and parameters for the model can be found in section 3.8.

The HK is activated via auto-phosphorylation to form $HK_p$, and the transfer of this phosphate group activates the RR to $RR_p$. $RR_p$ is inactivated either through phosphatase activity of the HK

or auto-dephosphorylation. Key reactions that describe this process in our model are:

$$\mathrm{HK} \underset{k^r_{\mathrm{HK_p}}}{\overset{k^f_{\mathrm{HK_p}}}{\rightleftharpoons}} \mathrm{HK_p}, \tag{3.1}$$

$$\mathrm{HK_p} + \mathrm{RR} \underset{k^r_{\mathrm{NCP}}}{\overset{k^f_{\mathrm{NCP}}}{\rightleftharpoons}} \mathrm{HK} + \mathrm{RR_p}, \tag{3.2}$$

$$\mathrm{RR_p} \xrightarrow{k_{\mathrm{dephos}}} \mathrm{RR}. \tag{3.3}$$

Since our design is dependent on activation of the RR with respect to the amount of reference protein (scaffold) that is available for binding, we expect the non-scaffold mediated rate of RR activation (described by reaction 3.2) to be low. Accordingly, rate constants were chosen to be similar to those between a HK and its non-cognate RR [40]. We assume that our reference protein has two independent binding sites to which a RR and HK can bind, and that this binding is independent of the phosphorylation state of HK and RR. When both $\mathrm{HK_p}$ and RR are bound to a scaffold, RR activation occurs at a faster rate, chosen to be similar to the rate of phosphorylation between a HK and its cognate RR. Some of the reactions that describe this process in our model are:

$$\mathrm{Sc} + \mathrm{HK_p} \underset{k^r_{\mathrm{SH_3}}}{\overset{k^f_{\mathrm{SH_3}}}{\rightleftharpoons}} \mathrm{ScHK_p}, \tag{3.4}$$

$$\mathrm{ScHK_p} + \mathrm{RR} \underset{k^r_{\mathrm{LZX}}}{\overset{k^f_{\mathrm{LZX}}}{\rightleftharpoons}} \mathrm{ScHK_pRR}, \tag{3.5}$$

$$\mathrm{ScHK_pRR} \underset{k^r_{\mathrm{CP}}}{\overset{k^f_{\mathrm{CP}}}{\rightleftharpoons}} \mathrm{ScHKRR_p}. \tag{3.6}$$

Activated RR, $\mathrm{RR_p}$ allows for the production of our target protein, AS with the reaction:

$$\mathrm{RR_p} \xrightarrow{\beta_{AS}} \mathrm{RR_p} + \mathrm{AS} \tag{3.7}$$

The creation of the anti-scaffold uses a Hill equation to describe its kinetic rate. We assume that the rate of production of AS from $\mathrm{RR_p}$ is unaffected by any other molecules $\mathrm{RR_p}$ is bound to. Therefore, the rate of production of AS is:

$$\frac{d[\mathrm{AS}]}{dt} = \beta_{AS}\left[\beta_0 + K_d\left(\frac{f}{g}\right)\right] \tag{3.8}$$

$$f = \mathrm{RR_p} + \mathrm{ScRR_p} + \mathrm{ScHKRR_p} + \mathrm{ScHK_pRR_p} + \mathrm{ScRR_pAS}$$

$$g = K_d + \mathrm{RR_p} + \mathrm{ScRR_p} + \mathrm{ScHKRR_p} + \mathrm{ScHK_pRR_p} + \mathrm{ScRR_pAS}$$

AS contains the same protein domains as RR and HK, enabling it to competitively bind to and sequester Sc. As with HK and RR binding to Sc, we assume that As can bind to either of these sites independently. However, once AS is bound to Sc, if an occupied binding site becomes available, we assume that AS is able to bind to and sequester this site very rapidly. Furthermore, we assume that once AS is bound to both of the sites in Sc, the sequestration of Sc by AS is irreversible. Some of the reactions that describe AS sequestration of Sc are:

$$\mathrm{Sc} + \mathrm{AS} \xrightarrow{k^f_{\mathrm{seq}}} \mathrm{ScAS} \tag{3.9}$$

$$\mathrm{ScHK} + \mathrm{AS} \underset{k^r_{\mathrm{LZX}}}{\overset{k^f_{\mathrm{LZX}}}{\rightleftharpoons}} \mathrm{ScHKAS} \xrightarrow{k^r_{\mathrm{SH_3}}} \mathrm{ScAS} + \mathrm{HK} \tag{3.10}$$

$$\mathrm{ScRR} + \mathrm{AS} \underset{k^r_{\mathrm{SH_3}}}{\overset{k^f_{\mathrm{SH_3}}}{\rightleftharpoons}} \mathrm{ScRRAS} \xrightarrow{k^r_{\mathrm{LZX}}} \mathrm{ScAS} + \mathrm{RR} \tag{3.11}$$

We have implemented our model using the SimBiology toolkit in MATLAB, which converts the reactions into a set of ODEs. Using the numerical solver ode23t, the simulations were run until each of the species reached steady-state. To evaluate the behavior of our circuit, simulations were performed over a range of steady-state Sc and RR concentrations. This was done by changing the rate of synthesis of the corresponding species. We compared the results of this simulation to the open loop circuit, which used the same model without the sequestration reactions. Figures 3.2 and 3.3 summarize the results of these simulations.

In the simulations, the addition of the negative feedback regulates the level of our target protein dependent on the amount of Sc in the system. The maximum reference level that our target is capable of tracking is dependent on the amount of RR in the system, which mediates the scaffold-

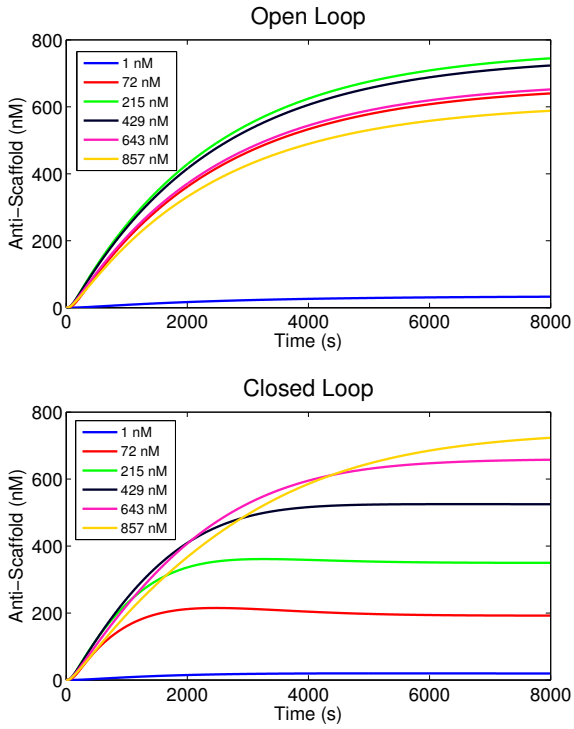



Figure 3.2: Time trace of total anti-scaffold (AS) concentration for varying scaffold(Sc) concentrations at a fixed response regulator concentration (350 nM).

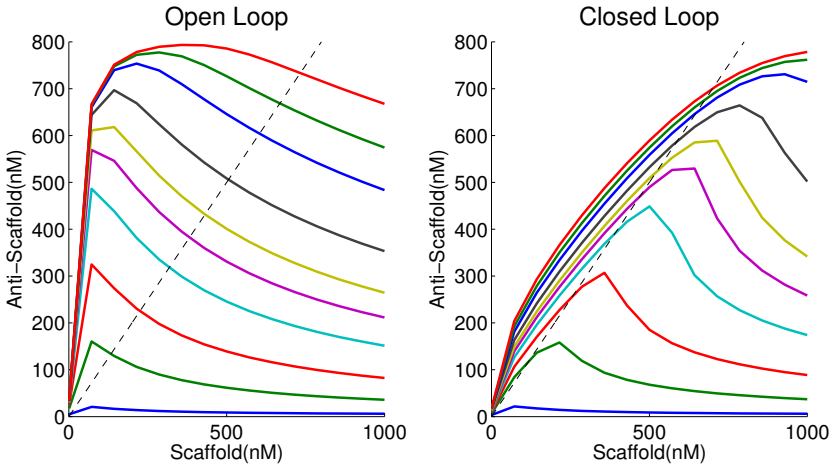



Figure 3.3: Steady-state simulations for different $Sc$ and $RR$ concentrations. The presence of the negative feedback loop is able to regulate the level of our target molecule, the anti-scaffold (AS), allowing it to track the level of our reference, the scaffold (Sc).

dependent expression of our target protein. As we increase the RR levels, we achieve the maximum amount of target protein that can be produced. The range in which the target protein is able to track the reference protein also increases. For a fixed RR concentration, the amount of target protein decreases from a maximum value as Sc increases. This is because co-localization of the HK and RR decreases if there is too much scaffold in the system. The regulation of the production of the target protein by the feedback loop is evident when compared to the open loop case where the target protein level quickly reaches its maximum concentration, and then decreases due to the loss of co-localization with increasing Sc.

In order to further characterize our circuit, and to determine which parameters our system were most sensitive to, we used the SimBiology toolkit to numerically calculate the sensitivity of our output, and the total target protein concentration ($x_{\mathrm{TOT}}$) with respect to the different parameters in our model. Specifically we calculated:

$$\frac{dx_{\mathrm{TOT}}}{x_{\mathrm{TOT}}} \Big/ \frac{d\theta_e}{\theta_e}$$

Figure 3.4 shows the normalized sensitivities of the total amount of target protein with respect to different parameters for a fixed scaffold concentration (572 nM) and increasing RR concentrations. We reasoned that increasing RR concentrations would be analogous to increasing the gain of our feedback circuit. Our circuit is most sensitive to parameters affecting the phosphorylation rates of the HK and RR, $k^f_{\mathrm{HK^P}}$,$k^r_{\mathrm{HK^P}}$, $k_{dephos}$, parameters dealing with sequestration, $k^f_{\mathrm{SH_3}}$, $k^r_{\mathrm{SH_3}}$, $k^r_{\mathrm{LZX}}$, $k^r_{\mathrm{LZX}}$, $k^r_{\mathrm{seq}}$, and parameters affecting the production of HK, Sc, and RR, $\beta_{Sc}$, $\beta_{RR}$, and $\beta_{HK}$. The sensitivity analysis informs us on which parameters have the most impact on circuit performance. By choosing to focus on parameters that are more easily tuned experimentally, we can utilize the model to optimize the circuit design process.

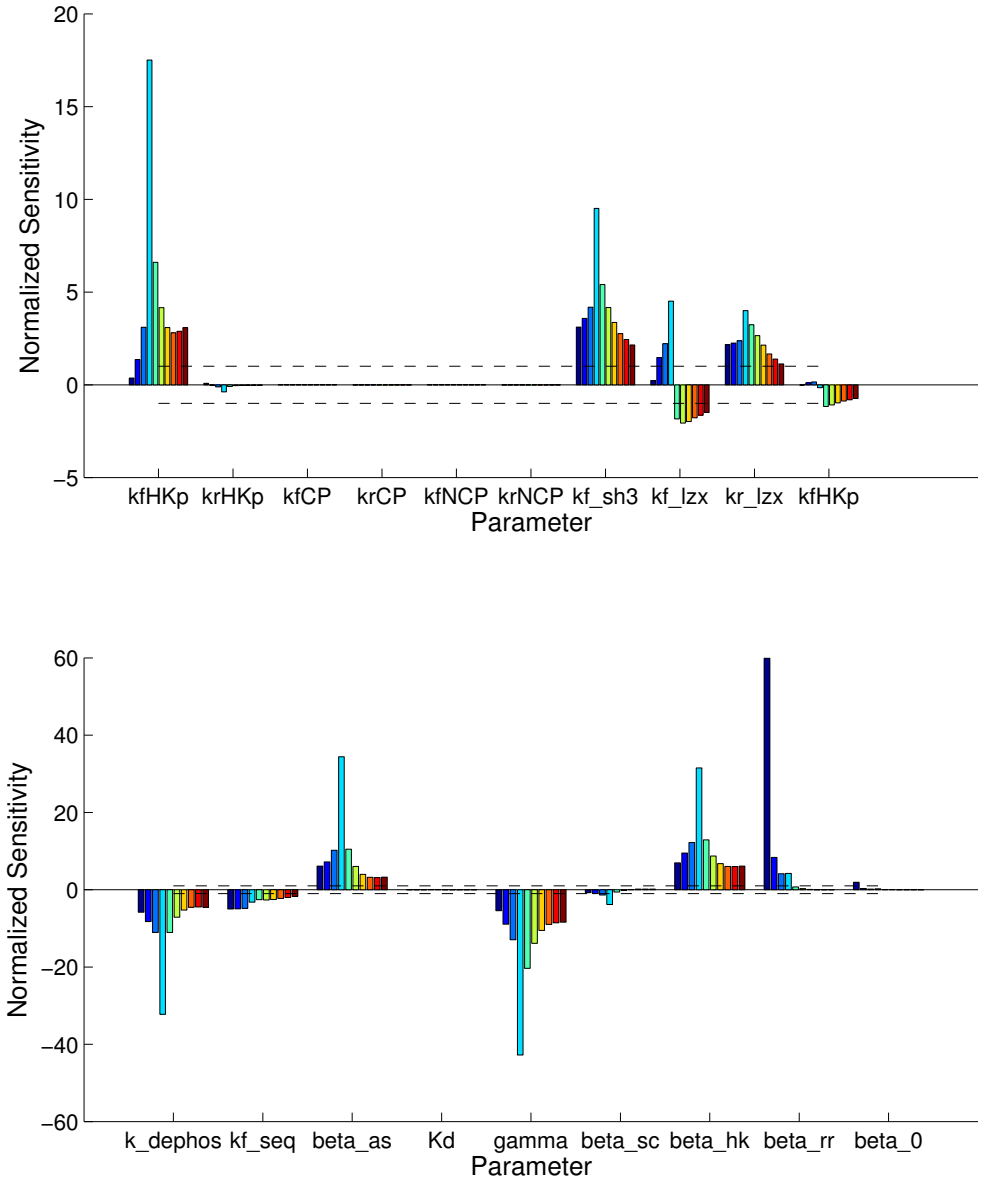



Figure 3.4: Sensitivity analysis showing the normalized sensitivity of total anti-scaffold to the different parameters in our model. The total amount of anti-scaffold is most dependent on the production and degradation rates of the various species, the affinity of the programmable scaffolds, and the deactivation rate of the response regulator. For the level of histidine kinase used in these simulations (2 nM), the total anti-scaffold concentration is invariant with the auto-activation rates of the histidine kinase. Colors indicate different levels of response regulator concentrations from the lowest value (1 nM, blue) to the highest value (500 nM, red) sampled.

## 3.5 Preliminary Experimental Results

Using plasmids obtained from WR Whitaker and JE Dueber, standard cloning practices were used to assemble the circuit onto three plasmids. The reference (scaffold) component was cloned into a high-copy plasmid under a tetracycline-inducible promoter. The histidine kinase (Taz) was cloned into a low-copy plasmid under a constitutive promoter. The response regulator (CusR) and target (anti-scaffold) were cloned into a medium-copy plasmid, with the response regulator under an arabinose-inducible promoter and the anti-scaffold under a response-regulator-activated promoter.

We implemented our circuit *in vivo* in a ΔCusS ΔCusR strain of *E. coli* [39]. This was done to reduce any non-scaffold mediated activation of the RR. In the absence of CusS, activated CusR proteins remain phosphorylated. This is because CusS serves as both the histidine kinase and the phosphatase in the native system.

Characterization of circuit behavior was done by independently varying the levels of Sc and RR. The inducible promoters allowed for the tunable expression of Sc and RR by anhydrous tetracycline (aTc) and arabinose, respectively. The Sc and AS were fused to fluorescent reporters, providing real-time circuit behavior data in bulk and single cell measurements. Cells were induced with 0–150 nM of anhydrous tetracycline (aTc) and 0 - 0.01% arabinose. A modified version of MOPS EZ Rich media (Teknova, M2105) is used to reduce the effect of media coloration on fluorescence measurements. Induced cultures are grown overnight (20 hrs, 37C) and bulk fluorescence measurements (GFP:488/520nm, RFP:580/610nm) are taken with a BioTek H1F microplate reader. An open loop version of the circuit, which contained a fluorescent reporter but none of the domains required to sequester the Sc, was tested in parallel with our closed loop circuit.

Overnight cultures for the open and closed loop circuits were grown overnight and steady state fluorescence values for the scaffold-mCherry, GFP (open loop), and anti-scaffold-GFP (closed loop) fusion proteins were measured. Figures 3.5–3.6 show the steady state values of scaffold versus anti-scaffold in terms of total normalized RFP versus GFP fluorescence.

In the open loop circuit (Figure 3.5), we see that with no response regulator (Ara 0%), GFP induction is close to zero. As response regulator concentration increases, the steady state level

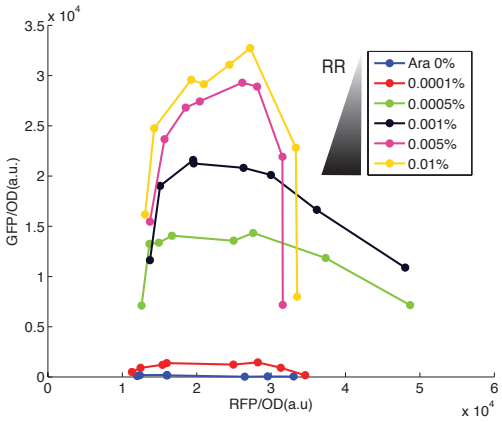



Figure 3.5: Steady state fluorescence values for the open loop circuit with scaffold-mCherry fluorescence on the x-axis (0 - 150nM aTc induction, L-R) and GFP values on the y-axis. Increasing arabinose concentrations indicate increasing response regulator concentrations.

of GFP peaks at some optimal concentration of scaffold (2–3 $\times$ $10^4$ a.u) and then decreases with increasing scaffold induction. We hypothesize that as scaffold concentrations exceed those of the response regulator and histidine kinase, each free scaffold will be bound to only one or the other, effectively reducing the amount of activated response regulators. This single-occupancy effect is one that has previously been shown [41]. We also see that with higher concentrations of response regulator, the scaffold seemed to saturate at a lower maximum concentration, resulting in the curves shown for 0.005% and 0.01% arabinose. Finally, the effect of leaky promoter activation should also be considered. In the case of no scaffold induction (0 nM aTc, first data point), we would expect to still see some background level of scaffold production, which, when combined with an excess of response regulator and histidine kinases to occupy those scaffolds, results in increasing background levels of anti-scaffold. Looking at the $x$-axis of Figure 3.5, we see that scaffold concentrations start at $1 \times 10^4$ a.u. with 0 nM aTc induction. As the arabinose induction of response regulator increases, so too does the steady state concentration of anti-scaffold.

In the closed loop circuit (Figure 3.6), we expect the steady state anti-scaffold concentrations to increase more gradually with increasing scaffold concentrations, which is apparent in the experimental data.

Rather than peaking at $2 \times 10^4$ a.u. scaffold fluorescence, the closed loop circuit asymptotes closer to $4 \times 10^4$ a.u. and does not yet show the effects of single scaffold occupancy. We believe

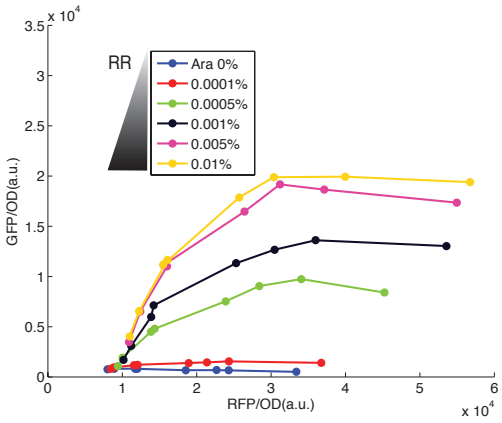



Figure 3.6: Steady state fluorescence values for the closed loop circuit with scaffold-mCherry fluorescence on the x-axis (0 - 150nM aTc induction, L-R) and anti-scaffold-GFP values on the y-axis. Increasing arabinose concentrations indicate increasing response regulator concentrations.

this is consistent with our model, since anti-scaffold sequestration effectively reduces the amount of free scaffold such that it actually lessens the probability of single-scaffold occupancy. Based on the model, we predict that if the scaffold induction range were increased beyond 150 nM, the steady state anti-scaffold concentration would show a similar decrease as the open loop, just at higher scaffold concentrations. Finally we see that although it has nearly the same level of leaky scaffold production, the anti-scaffold levels are much lower than their open-loop counterparts.

## 3.6 Modifications to the Model and Experimental Results

As we characterized the circuit more, we observed that we were seeing higher than expected activation of proteins downstream of the CusR promoter. We noted that our circuit was implemented in a ΔCusS ΔCusR *E. coli* knockout strain [45]. In the absence of CusS, the native bifunctional histidine kinase/phosphatase partner for CusR, activated CusR proteins remain phosphorylated. Accordingly, we re-introduced a CusS(G448A) mutant behind an inducible promoter to tune response regulator deactivation. The G448A mutation disrupts the ATP binding site, eliminating kinase auto-phosphorylation without affecting phosphatase activity [46, 47]. We modified the model to reflect this change by changing reaction 3.3 to

$$\mathrm{RR_p} + \mathrm{Ph} \underset{k^r_{\mathrm{Ph}}}{\overset{k^f_{\mathrm{Ph}}}{\rightleftharpoons}} \mathrm{RR_p{:}\ Ph} \xrightarrow{k_{\mathrm{dephos}}} \mathrm{RR} + \mathrm{Ph}$$

Since the activated response regulator dimerizes when it serves as a transcription factor, we changed the hill coefficient in the anti-scaffold synthesis equation to 2. Reaction (3.8) was modified to reflect this:

$$\frac{d[\mathrm{AS}]}{dt} = \beta_{AS}\left[\beta_0 + \left(\frac{f^2}{K_D^2 + f^2}\right)\right]$$

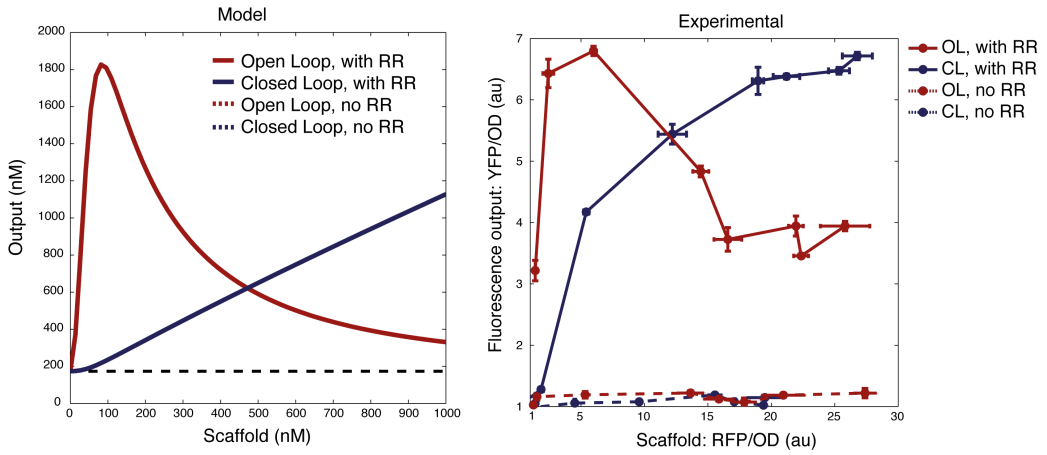
$$f = \mathrm{RR_p} + \mathrm{ScRR_p} + \mathrm{ScHKRR_p} + \mathrm{ScHK_pRR_p} + \mathrm{ScRR_pAS}$$

Dephosphorylation is now modeled enzymatically as a two-step reaction where the phosphatase binds to the phosphorylated response regulator first and then the phosphate is removed irreversibly, in a Michalis-Menten reaction. This is consistent with previous models [48].

The final model consists of 80 reactions, 25 differential equations, and 26 parameters. Many parameters were selected from experimental values found in the literature [43, 49, 44], and others were estimated within a physiologically reasonable range.

Experimental data recapitulated the model predictions for the circuit in steady-state conditions (Figure 3.7). The circuit was also able to respond dynamically to increases in aTc concentration, closely following the predictions of the model (Figure 3.8). Interestingly, a model based exploration of parameter space predicted that the ratio between the scaffold and the anti-scaffold could be tuned by varying the phosphatase and response regulator concentrations (Figure 3.9). This corresponded to tunable knobs, we could alter the response regulator concentration experimentally by varying the arabinose concentration. Experimental results suggested that this was the case using YFP/RFP as a proxy for this ratio (Figure 3.10). The final circuit not only serves as a biomolecular concentration tracker, but it provides a mechanism to maintain two proteins at a tunable fixed ratio dynamically

with changing environmental conditions.



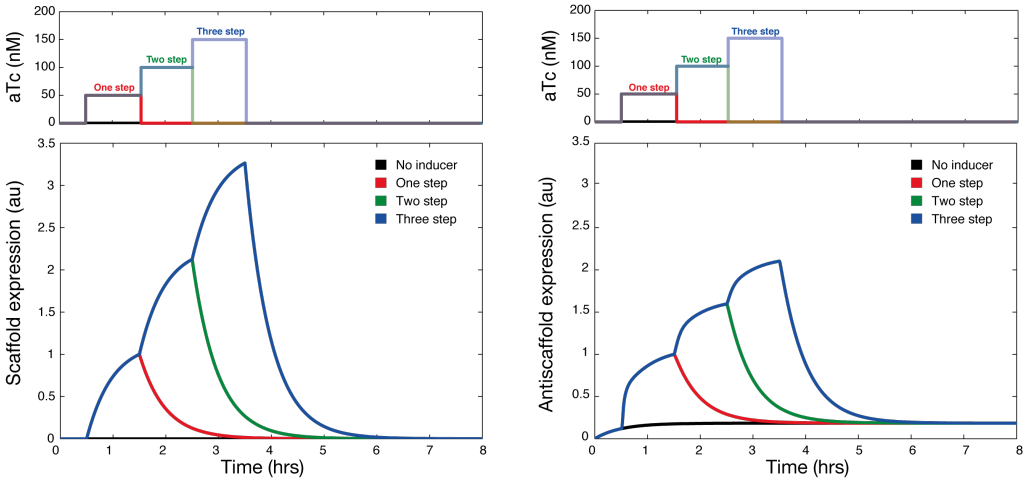

(a) (b)

Figure 3.7: **Open loop versus closed loop.** (a) Model predictions of scaffold circuit with and without negative feedback. Solid lines show anti-scaffold output over a range of scaffold concentrations (0 - 1000 nM) for open and closed loop circuits with constant response regulator (100 nM). Dotted lines show lack of output in the absence of response regulator. Open loop circuit shows scaffold single occupancy effect at lower levels of scaffold. (b) Steady state experimental data of open and closed loop circuits with and without response regulator matches model predictions. Both sets of experimental data were normalized by the autofluorescence of a control *E. coli* strain. *Figure and caption obtained from [35].*

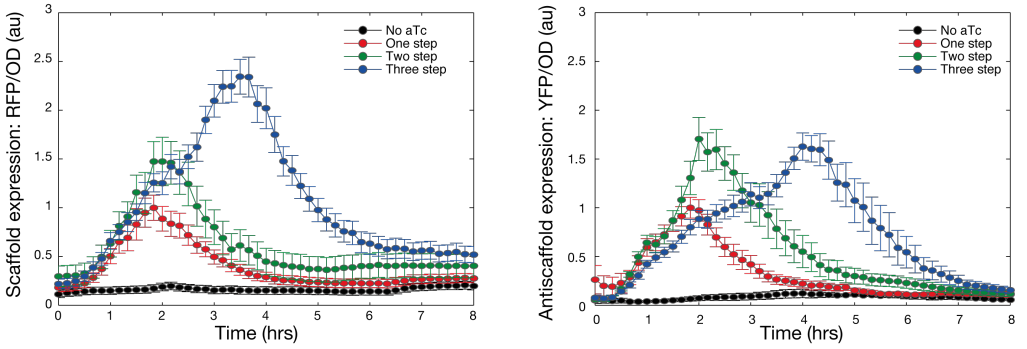
## 3.7 Conclusion

In this chapter, we have mathematically modeled and shown experimental results for a biomolecular concentration tracker. We believe that this is a promising mechanism for regulating the level of protein in the cells based on changing conditions. By linking the level of the reference protein to a signal of interest, you can dynamically affect the levels of downstream proteins, or keep the ratio between proteins fixed. We believe that allowing proteins to maintain a fixed ratio, tunable by different components in the circuit, is particularly useful in synthetic metabolic pathways, where altering the ratios of enzymes from different sources to balance flux can yield an increase in overall yield [38]. This circuit enables that at the protein level and can potentially replace promoter tuning for synthetic circuits. Promoter tuning balances circuit component levels more indirectly, since it operates at the DNA level, and dynamic response is also slower and more difficult to implement.

We note that the circuit presented here can serve as a template for the dynamically maintaining proteins at fixed ratios and will be a valuable addition to the synthetic biology toolkit.




(a)




(b)

Figure 3.8: **Multi-step induction of tracker circuit.** (a) Simulation results for a three step induction show overlapping response times with each curve decreasing based on degradation rate after induction ceases. Upper panel shows aTc induction pattern with one hour steps increasing in 50 nM increments starting 30 minutes after start of experiment. (b) Experimental time traces for Sc-RFP show overlapping fluorescence output, with each curve decreasing at a time proportional to the number of steps. Corresponding anti-scaffold-YFP data show similar overlaps and proportional decreases. Fluorescent measurements are normalized such that the maximum value of the one step curve is 1 a.u. to better visualize fold change. *Figure and caption obtained from [35].*

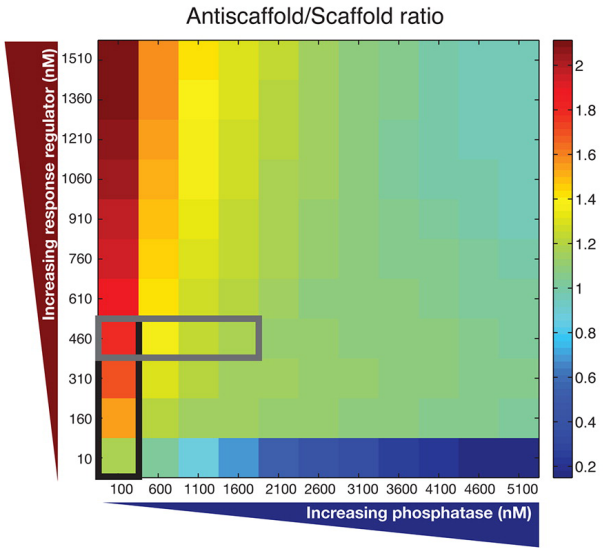
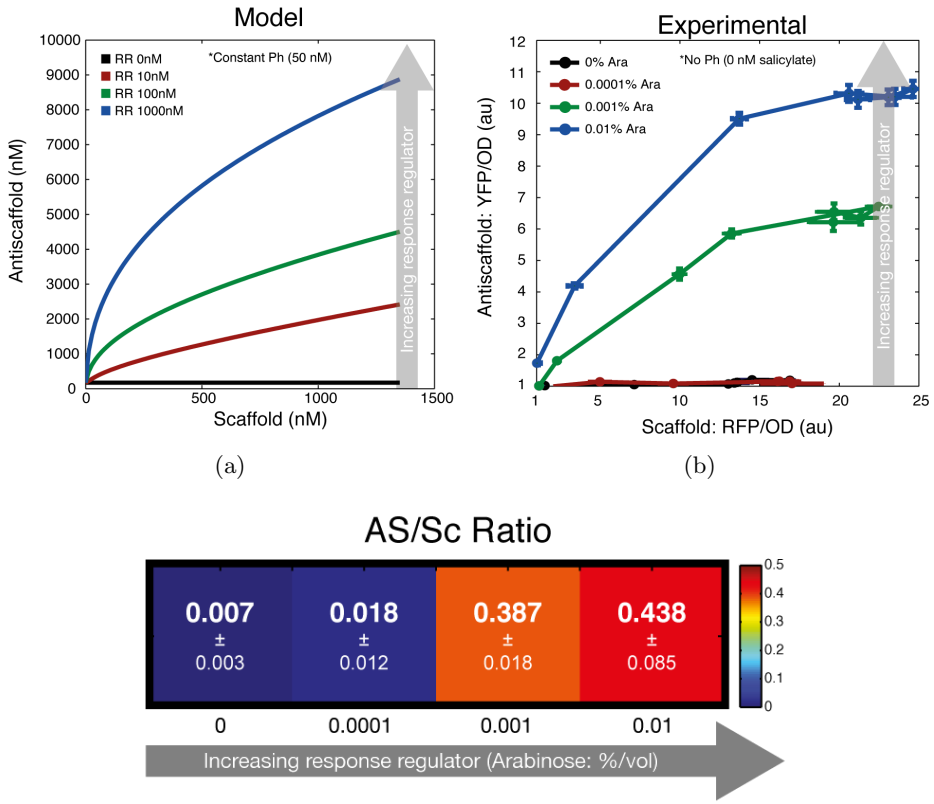



Figure 3.9: **Model predicted tunability of anti-scaffold to scaffold ratio.** Heat map showing anti-scaffold to scaffold ratio. Increasing response regulator results in greater AS/Sc ratios. Gray box represents estimated experimental phosphatase induction range. Black box estimates experimental response regulator induction range. *Figure and caption adapted from [35].*




(a) (b)

(c)

Figure 3.10: **Steady state experimental tuning of response regulator.** (a) Simulation data of input-output curves with increasing response regulator concentrations (01000 nM). Increasing response regulator increases the scaffold occupancy limit as well as overall AS/Sc ratio. (b) Experimental data of steady state scaffold to anti-scaffold curves with 10-fold increases in response regulator induction (00.01% arabinose). (c) Ratios of YFP/RFP from part B as a proxy for As/Sc ratios with increasing response regulator.*Figure and caption adapted from [35].*

## 3.8 Model Reactions and Parameters

### 3.8.1 Chemical Reactions

Production and Degradation Reactions (other species are degraded with rate $\gamma$ and not produced):

$$\phi \underset{\gamma}{\overset{\beta_{HK}}{\rightleftharpoons}} HK \tag{3.12}$$

$$\phi \underset{\gamma}{\overset{\beta_{RR}}{\rightleftharpoons}} RR \tag{3.13}$$

$$\phi \underset{\gamma}{\overset{\beta_{Sc}}{\rightleftharpoons}} Sc \tag{3.14}$$

Auto-phosphorylation of Histidine Kinase:

$$HK \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} HK_p \tag{3.15}$$

$$ScHK \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} ScHK_p \tag{3.16}$$

$$ScHKRR \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} ScHK_pRR \tag{3.17}$$

Non-scaffold mediated background phosphorylation:

$$HK_p + RR \underset{k_{rNonCog}}{\overset{k_{fNonCog}}{\rightleftharpoons}} HK + RR_p \tag{3.18}$$

Scaffold binding to HK (SH3) or RR (LZX):

$$Sc + HK \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHK \tag{3.19}$$

$$Sc + HK_p \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHK_p \tag{3.20}$$

$$Sc + RR \underset{k_{rScRR}}{\overset{k_{fScRR}}{\rightleftharpoons}} ScRR \tag{3.21}$$

$$Sc + RR_p \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScRR_p \tag{3.22}$$

Trimer formation:

$$ScHK + RR \underset{k_{rScRR}}{\overset{k_{fScRR}}{\rightleftharpoons}} ScHKRR \tag{3.23}$$

$$ScHK_p + RR \underset{k_{rScRR}}{\overset{k_{fScRR}}{\rightleftharpoons}} ScHK_pRR \tag{3.24}$$

$$ScHK + RR_p \underset{k_{rScRR}}{\overset{k_{fScRR}}{\rightleftharpoons}} ScHKRR_p \tag{3.25}$$

$$ScHK_p + RR + p \underset{k_{rScRR}}{\overset{k_{fScRR}}{\rightleftharpoons}} ScHK_pRR_p \tag{3.26}$$

$$ScRR + HK \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHKRR \tag{3.27}$$

$$ScRR + HK_p \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHK_pRR \tag{3.28}$$

$$ScRR_p + HK \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHKRR_p \tag{3.29}$$

$$ScRR_p + HK_p \underset{k_{rScHK}}{\overset{k_{fScHK}}{\rightleftharpoons}} ScHK_pRR_p \tag{3.30}$$

Phosphorylation and Auto-dephosphorylation of Response Regulator (in *ACS Synthetic Biology* paper, these reactions are replaced with enzymatic dephosphorylation):

$$ScHK_pRR \underset{k_{rCogP}}{\overset{k_{fCogP}}{\rightleftharpoons}} ScHKRR_p \tag{3.31}$$

$$RR_p \xrightarrow{k_{AutoDephos}} RR \tag{3.32}$$

$$ScRR_p \xrightarrow{k_{AutoDephos}} ScRR \tag{3.33}$$

$$ScHKRR_p \xrightarrow{k_{AutoDephos}} ScHKRR \tag{3.34}$$

$$ScHK_pRR_p \xrightarrow{k_{AutoDephos}} ScHK_pRR \tag{3.35}$$

$$ScRR_pAS \xrightarrow{k_{AutoDephos}} ScRRAS \tag{3.36}$$

Sequestration of Scaffold by Anti-scaffold:

$$Sc + AS \xrightarrow{k_{fScAS}} ScAS \tag{3.37}$$

$$ScHK + AS \underset{k_{rAS_{LZX}}}{\overset{k_{fAS_{LZX}}}{\rightleftharpoons}} ScHKAS \xrightarrow{k_{rAS_{SH3}}} ScAS + HK \tag{3.38}$$

$$ScHK_p + AS \underset{k_{rAS_{LZX}}}{\overset{k_{fAS_{LZX}}}{\rightleftharpoons}} ScHK_pAS \xrightarrow{k_{rAS_{SH3}}} ScAS + HK_p \tag{3.39}$$

$$ScRR + AS \underset{k_{rAS_{SH3}}}{\overset{k_{fAS_{SH3}}}{\rightleftharpoons}} ScRRAS \xrightarrow{k_{rAS_{LZX}}} ScAS + RR \tag{3.40}$$

$$ScRR_p + AS \underset{k_{rAS_{SH3}}}{\overset{k_{fAS_{SH3}}}{\rightleftharpoons}} ScRR_pAS \xrightarrow{k_{rAS_{LZX}}} ScAS + RR_p \tag{3.41}$$

Synthesis of Anti-Scaffold:

$$\phi \underset{\gamma}{\overset{k_f}{\rightleftharpoons}} As$$

### 3.8.2 Parameters

| **Parameter** | **Value** | **Units** | **Description** |
|---|---|---|---|
| $\beta_{AS}$ | 10 | nM $s^{-1}$ | anti-scaffold production from RRp |
| $\beta_{HK}$, $\beta_{RR}$, $\beta_{sc}$ | [steady-state value]/$\gamma$ | nM $s^{-1}$ | these were constants and set as input values for the model |
| $\beta_0$ | 0.01 | | leaky anti-scaffold production (1% of total induction) |
| $\gamma$ | 3.84 x $10^{-4}$ | $s^{-1}$ | degradation and dilution [49, 50] |
| $K_D$ | 100 | nM | constant in anti-scaffold activation Hill Function |

| **Forward and Reverse Reaction Rates** | | | | |
|---|---|---|---|---|
| **Parameter** | | **Value** | **Unit** | **Description** |
| kHKp | kf | 0.003 | $s^{-1}$ | HK auto-phosphorylation [49] |
| | kr | 0.0001 | $s^{-1}$ | [43] |
| kcogp | kf | 102.1 | $s^{-1}$ | cognate HK-RR phosphorylation [49] |
| | kr | 0.00294 | $s^{-1}$ | [49] |
| knoncog | kf | 0.0031 | $M^{-1}$ $s^{-1}$ | non-cognate HK-RR phosphorylation [49] |
| | kr | 0.0002 | $M^{-1}$ $s^{-1}$ | [49] |
| kSH3 | kf | 1 x $10^5$ | $M^{-1}$ $s^{-1}$ | SH3 domain/ligand binding [44] |
| | kr | kf:SH3 (0.1 x $10^{-6}$) | $s^{-1}$ | $K_D = 0.1M$[41] |
| kLZX | kf | 1 x $10^5$ | $M^{-1}$ $s^{-1}$ | leucine zipper binding |
| | kr | kf:LZX (0.01 x $10^{-6}$) | $s^{-1}$ | $K_D = 0.01M$[41] |
| kSc:HK | kf | 4kf:SH3 | $M^{-1}$ $s^{-1}$ | scaffold binding to HK with 4 SH3 domains |
| | kr | kr:SH3 | $s^{-1}$ | |
| kSc:RR | kf | kf:LZX | $M^{-1}$ $s^{-1}$ | scaffold binding to RR with 1 LZX domain |
| | kr | kr:LZX | $s^{-1}$ | |
| kSc:AS | kf | kf:LZX + kf:SH3 | $M^{-1}$ $s^{-1}$ | scaffold binding to anti-scaffold |
| | kr | 0.001kr:LZX | $s^{-1}$ | |
| kAS-SH3 | kf | kf:SH3 | $s^{-1}$ | anti-scaffold binding to Sc:RR complex |
| | kr | 0.001kr:SH3 | $s^{-1}$ | |
| kAS-LZX | kf | kf:LZX | $M^{-1}$ $s^{-1}$ | anti-scaffold binding to Sc:HK complex |
| | kr | 0.001kr:LZX | $s^{-1}$ | |

# Chapter 4

# Design of a Workflow for *De Novo* Computational Enzyme Design

## 4.1 Introduction

One of the big promises of computational protein design (CPD) is its potential for engineering designer enzymes. Enzymes are truly amazing catalysts, providing rate accelerations of up to $10^{19}$ over uncatalyzed reactions with great regio- and stereo-selectivity. Being able to harness this potential in a general way will provide a great boon by reducing the need for multistep organic syntheses with protecting groups and hazardous solvents. From a synthetic biology perspective, a general framework to build designer enzymes is potentially one of the biggest contributions that CPD can provide to augment the synthetic toolkit. Custom-built enzymes in synthetic pathways can streamline metabolic engineering of novel products and allow us to program cells with new functions that nature has not previously accessed.

Recent advances [51, 52, 53, 54] in *de novo* enzyme design have bolstered our confidence that designer enzymes will one day become a reality. However, much work needs to be done in order for this to be accomplished. Rate accelerations from computationally designed enzymes are several orders of magnitude below natural enzymes. Furthermore, attempts at computational design of certain, well-characterized reactions, such as the chorismate mutase reaction, have so far proven to

be unsuccessful.

This chapter describes a mostly automated workflow we developed to find suitable scaffolds for the chorismate mutase reaction with the goal of engineering chorismate mutase activity *de novo* into an inactive protein scaffold (Figure 4.1). The workflow begins with crystal structures from the PDB that meet a set of criteria for scaffolds for computational protein design. Structures that had ligands that were similar in size to our target substrate and contained a wild-type arginine, a key amino acid for the chorismate mutase reaction, were selected for targeted ligand placement. After targeted ligand placement, designs were down-selected by visual inspection of the computational models of the active site. Molecular dynamics was then performed to look for promising candidates. Candidates that passed certain criteria from the molecular dynamics screen were synthesized and characterized experimentally. Results from experimental characterization can then be used to guide more *in silico* exploration of the design space. We believe that this workflow can serve as a template for future enzyme design projects alternating iteratively between computational screening and experimental characterization until we obtain an enzyme with the desired properties.

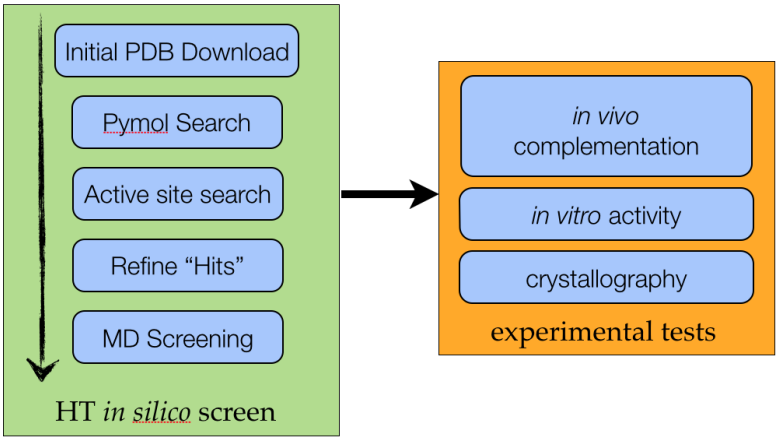



Figure 4.1: **Workflow for Enzyme Design.** The workflow leverages advances in computational protein design to focus a search of a large space of design possibilities into smaller areas that are experimentally tractable. Results from the experiments can then inform computational design algorithms to provide suggestions on other areas of design space to focus on, or areas to apply more computational power on to explore the space even more.

The work described in this chapter was a collaboration between myself, Timothy Wannier, and Bernardo Sosa Padilla Araujo. I conceptualized designing a framework to automate enzyme design.

Timothy and I wrote the scripts and performed the experiments. Preliminary molecular dynamics work was done by Bernardo, which I then continued adapting his scripts and protocol.

## 4.2 Background

Chorismate mutase catalyzes the pericyclic rearrangement of chorismate to prephenate (Figure 4.2A). It is one of the steps in the phenylalanine and tyrosine biosynthetic pathways. While crucial to central metabolism in lower organisms, this reaction is absent in mammals, making it an attractive target for antibiotics, fungicides, and herbicides. Because of this, the mechanism and transition state of the catalyzed reaction has been studied extensively [55, 56]. These studies have shown that enzymatic catalysis is achieved by stabilizing the chair-like transition state through electrostatic interactions with sidechains. Chorismate mutase seemed like a classic target for *de novo* computational enzyme design because of its straightforward mechanism—one-step reaction where the transition state was stabilized through electrostatic interactions, and because of *ab initio* calculations done to predict its transition state by the Houk Lab [56]. In fact, chorismate mutase was one of the enzymes that was used as a benchmark for the targeted ligand placement method developed at the Mayo lab [23]. In the study, the Phoenix Match algorithm (Targeted Ligand Placement) [23] was used to recapitulate enzyme active sites, including that of chorismate mutase. Phoenix Match was able to place an *ab initio* transition-state structure into the active site of chorismate mutase with the proper wild-type contacts. However, in spite of this success in recapitulating the active site in a wild-type scaffold, to the best of our knowledge, attempts to engineering chorismate mutase activity into a non-native scaffold have been unsuccessful.

In retrospect, engineering chorismate mutase activity is fairly challenging because of the multiple catalytic contacts that are required. An ideal chorismate mutase active site would preferentially bind the diaxial form of chorismate, as it is closer to the transition state (Figure 4.2B). This requires seven catalytic contacts including two buried arginines (Figure 4.2C), making targeted ligand placement into a new scaffold difficult. However, recent advances in enzyme design and increased computational resources allowing for large-scale targeted ligand placement and the use of molecular dynamics (MD)

as a prescreen [53, 54, 57] have encouraged us to revisit this problem. We attempted to use the increase in computational resources to do a broad, automated, systematic survey of the design space to achieve the *de novo* engineering of a chorismate mutase.

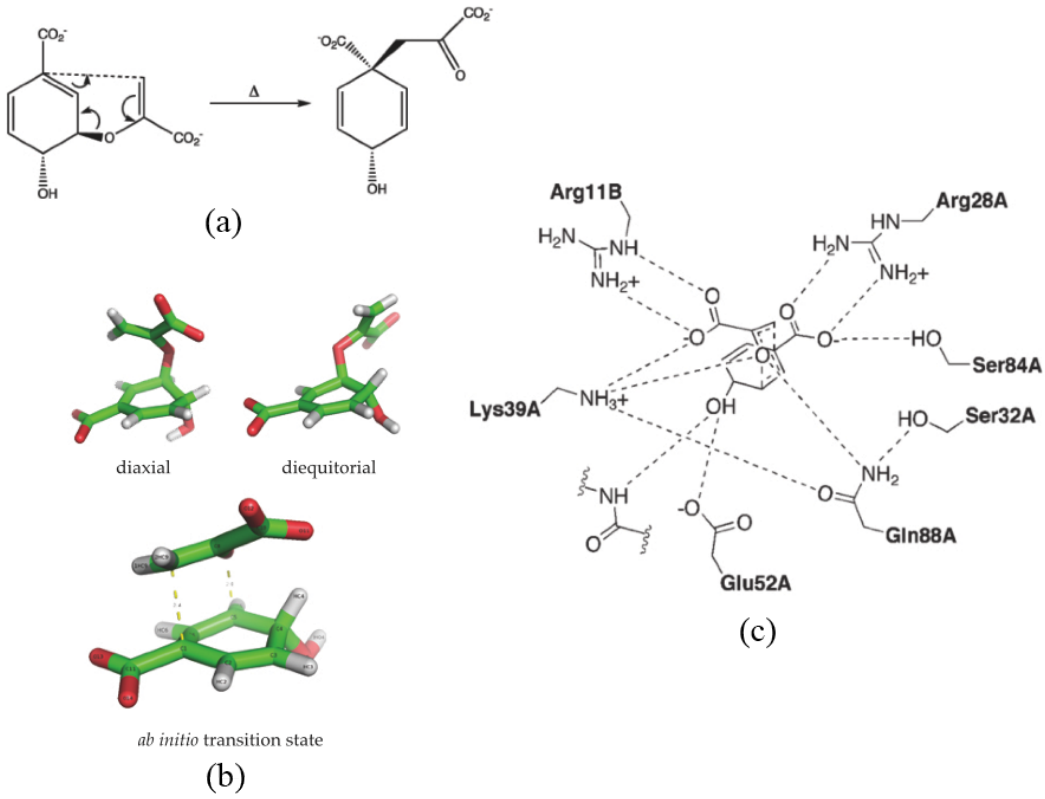



Figure 4.2: **Chorismate mutase reaction.** **(A)** Chorismate mutase catalyzes the conversion of chorismate to prephenate, the first committed step on the aromatic amino acid synthesis pathways. **(B)** Different conformations of chorismate; because the transition state is closer to the diaxial conformation, the enzyme stabilizes the diaxial form. **(C)** Active site of E. coli chorismate mutase. Multiple contacts bind preferentially to the diaxial form of chorismate, which helps stabilize the transition state. *Figures 1A and 1C were adapted from [23] Figure 1B was generated in PYMOL.*

## 4.3 Computational Enzyme Design Workflow for Chorismate Mutase

### 4.3.1 Finding Suitable Scaffolds from the Protein Data Bank

One of the challenges in enzyme design is finding appropriate scaffolds for Phoenix Match. This is especially true in the case of chorismate mutase, as it has proven difficult to find scaffolds with large pockets that contain at least one core arginine. We decided to automate the scaffold search process by starting with a set of structures from the Protein Data Bank that could be used for future active site searches.

We constructed a set of criteria (Table 4.1), taking into consideration features we usually look for when picking scaffolds for CPD such as resolution and expression organism. We downloaded 7,561 structures matching our criteria from the PDB. This set of structures is not specific to the chorismate mutase reaction and can be used for future scaffold searches.

Table 4.1: **Criteria for PDB Download.** List of criteria that used to filter the initial download from the Protein Data Bank.

| **Criterion** | **Value** |
|---|---|
| Expression Organism | *Escherichia coli* |
| "Has ligands" | Yes |
| X-ray Resolution | 0.1 Å–2.8 Å |
| Chain Length | 1–500 residues |
| Contains Protein | Yes |
| Contains DNA | No |
| Contains RNA | No |
| Contains DNA/RNA Hybrid | No |
| Number of Entities | 1 |
| Number of Chains | 1–2 |
| Remove similar sequences at 90% identity | |

To find potential scaffolds for chorismate mutase, we wrote a PYMOL script that analyzed the downloaded structures. The script first searched for structures that contained ligands with a similar radius of gyration to the chorismate mutase transition state (2.63 Å). Specifically, we searched for structures that contained ligands that had a radius of gyration between 1.95 and 4 Å. We defined the residues in the ligand binding pocket to be residues within a 4.5 Å shell of the ligand. We filtered the

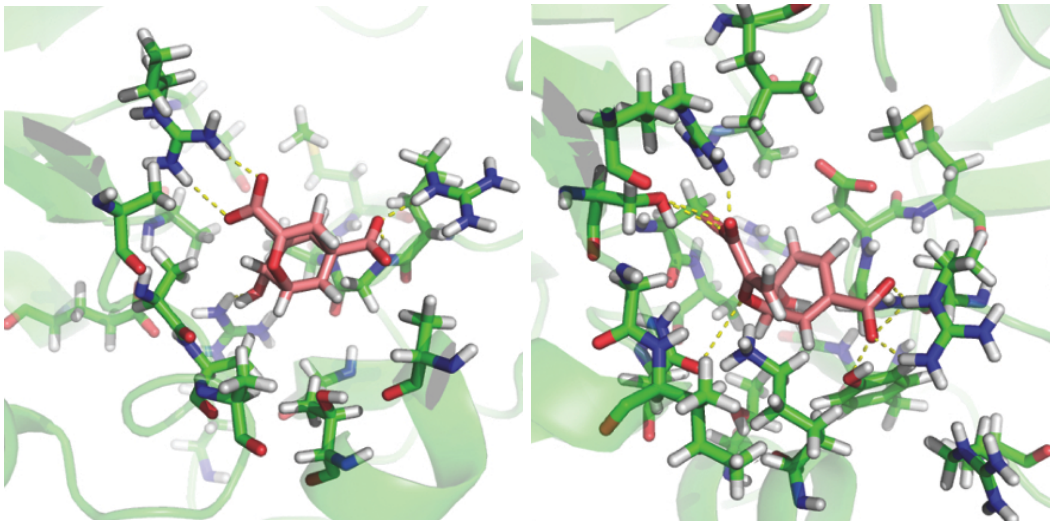
list of structures to those that contained a buried arginine (less than 15 $Å^2$ surface area exposed) in the ligand-binding pocket. Applying this filter to our list of scaffolds returned 451 potential scaffolds for the chorismate mutase reaction.

### 4.3.2 Active Site Search

We ran each of the scaffolds through proteinProcess, a Triad structure preparation script that adds hydrogens, builds missing residues and atoms, and resolves structural ambiguity for asparagine, glutamine, and histidine using a specified energy function (Biograf [58]). Most of the structures (429 of 451) passed through proteinProcess. Reasons for failing varied from too many missing residues to multiple atom densities. Phoenix Match was then performed on the remaining scaffolds using geometry definitions described in Lasilla *et al.* [23] using both backbone dependent (resolution 1.0 Å) and backbone independent (resolution 1.8 Å) conformer libraries [23]. We considered transition state placement to be successful if Phoenix Match returned a result that contained carboxyl interactions with 2 arginines (Figure 4.3A). The active-site search with the backbone-dependent library yielded 239 successful ligand placements, and the backbone-independent conformer library yielded 246 successes.

### 4.3.3 Refining the Active Site Search Hits

For each of the successful active sites, standard computational protein design was performed in Phoenix. Because of the smaller number of structures to analyze, we were able to use more computational resources per structure. This allowed us to greater explore paramater space. Specifically, we allowed for small rotations and translations of the ligand and used a finer resolution confomer library (1.0 Å for backbone-independent library and 0.1 Å for backbone-dependent library). In addition to the regular 100 kcal/mol energy bias that is standard in Phoenix Match, we included an energy bias of 50.0 kcal/mol for existing contacts or for the wild-type residue. The purpose of this bias was to minimize the number of mutations to the starting structure. There was no energy penalty for amino acids that did not make any of the required contacts. We analyzed the decoys outputted



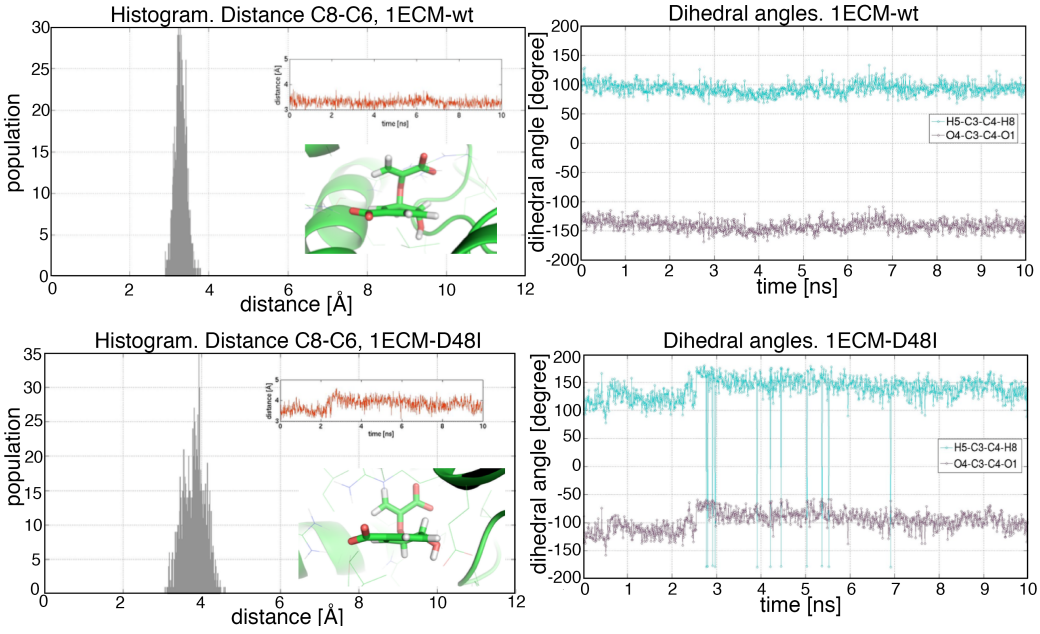
(a) (b)

Figure 4.3: **Sample active site search and refinement.** Active site searches were done on 429 PDB structures using geometry definitions modified from Lasilla *et al.* [23] with both backbone-dependent and -independent conformer libraries. **(a)** An active site search was considered successful if it had two arginine contacts to carboxyl groups. This is a sample of a successful active site search (scaffold PDB ID: 2RJH). **(b)** Successful active site searches were refined with a higher resolution rotamer library, allowing for small rotations and translations in the substrate. Shown here is the result following refinement of the active site search from Figure 2A. Extra catalytic contacts from a lysine and tyrosine are now present. Successful refinements (2 arginine contacts and a contact to the ether oxygen) were examined by eye and structures were chosen for MD simulations.

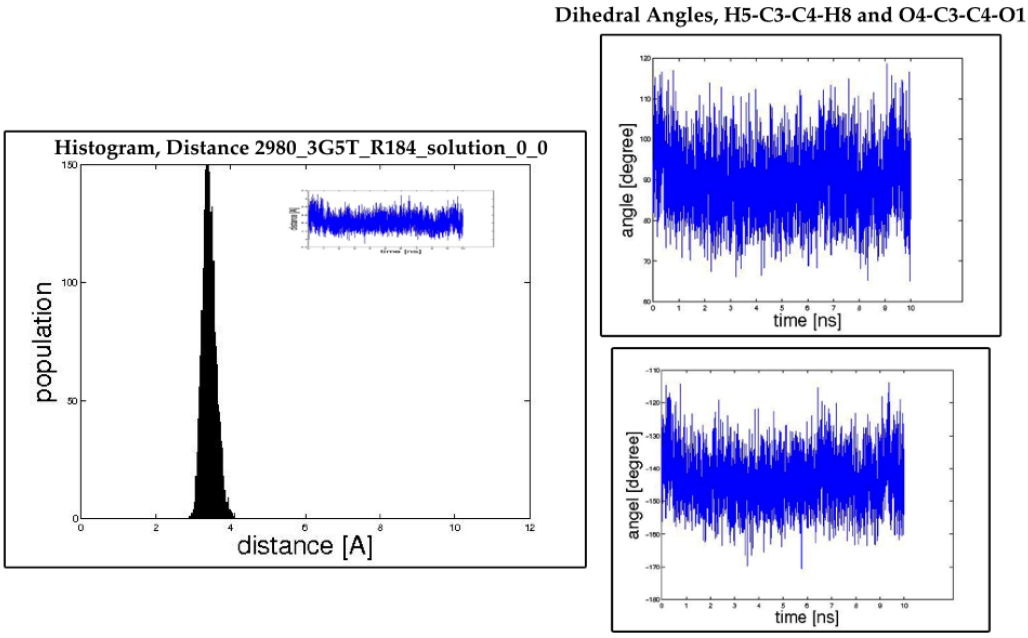
by the algorithm and evaluated designs to select candidates for further analysis. Specifically, we chose results that retained the dual arginine contacts and had at least one hydrogen bond interaction to the ether oxygen (which was found to be a key contact in previous mechanism studies) as candidates for further analysis (Figure 4.3B). Repacking with the backbone-dependent library returned 68 candidate structures, and repacking with the backbone-independent library returned 91 candidate structures. We visually inspected each of these structures and removed structures with unusual rotamers for catalytic contacts and whose designed active sites were too solvent exposed. After this analysis, 121 decoys were chosen for further screening with molecular dynamics.

### 4.3.4 Molecular Dynamics Screening

We performed molecular dynamics (MD) simulations on 121 candidate designs. Previous results have demonstrated the ability of MD simulations to predict the activity of *E. coli* chorismate mutase mutants that were previously generated in our lab (Figure 4.4A) [23, 57]. We used a similar molecular dynamics protocol to that applied to the *E. coli* chorismate mutase mutants, the MD simulations were run using the AMBER12 forcefield in an isothermal-isobaric (NPT) ensemble. The MD protocol consisted of three steps. First, the decoys went through 1000 steps of minimization under the AMBER forcefield. This was followed by a 10 picosecond equilibration step where the molecule and substrate started with harmonic energy constraints that were released as the simulation progressed. Finally, an unconstrained 10 nanosecond MD simulation was run. Through the course of the simulation, we monitored the distance between the C6 and C8 carbons, the two carbons where a new bond is formed. We also monitored the dihedral angles related to this bond, in order to give us an indication of whether or not the substrate stayed in the diaxial conformation, or moved to the diequatorial conformation (Figure 4.4B). Using this to analyze the decoys, we selected 10 hits from the 121 starting potential enzyme designs for experimental characterization (Table 4.2).




(a)




(b)

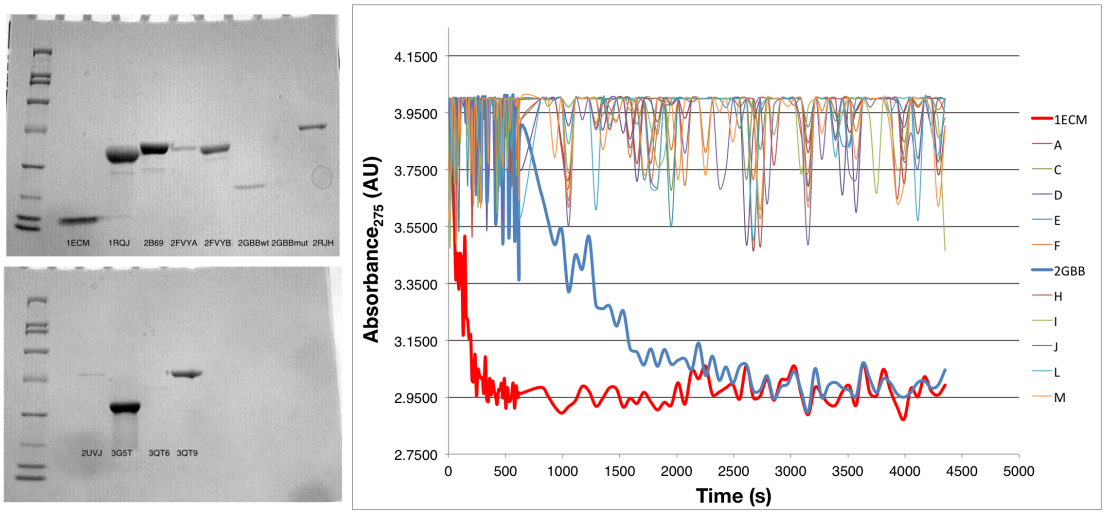
Figure 4.4: **Using molecular dynamics as a prescreen. (a)** Molecular dynamics simulations on *E. coli* chorismate mutase demonstrated its potential to be used as a pre-screen for enzyme design [57]. The protocol included looking at the stability of the C8-C6 distance (a bond forms here during the Claisen rearrangement) and the dihedral angles to monitor the diaxial and diequatorial states. A histogram and time-trace of the C8-C6 distance between the wild-type enzyme (top) and a less functional mutant (bottom) are shown. The wild-type enzyme has a tight histogram centered under 4 Å while the less functional mutant has a broader histogram with a greater C6-C8 distance. *Figure adapted from [59].* **(b)** Example of an MD hit. The C8-C6 distance and the dihedral angles appear to be stable throughout the entire 10 ns simulation.

Table 4.2: Molecular Dynamics Hits

| Scaffold PDB ID | Mutations from Wild-Type |
|---|---|
| 1RQJ | K46S, R97N, T183R |
| 2B69 | V191G, N194R, R215K, I249D, D275N |
| 2FVY | Y34G, F40G, M41R, D260K, N280G |
| 2FVY | Y34G, F40K. M41R, K116N, N280S |
| 2GBB | Q67R, I97N, Q104K |
| 2RJH | V33S, K35S, Y39N, Y344K |
| 2UVJ | W36S, W68S, K306R |
| 3G5T | Y14A, Y150K, Y242S, W289S |
| 3QT6 | W20N, S108K, M244R |
| 3QT9 | W62K, E129S, E393R, F408N |

### 4.3.5 Experimental Characterization of Hits from Molecular Dynamics Screen

We assembled genes encoding the MD hits using overlap extension PCR. As a control, we also assembled genes encoding the wild-type sequence of each of the scaffolds used. We evaluated the potential activity of each of the hits with an *in vivo* complementation assay [60] and an *in vitro* plate reader assay.

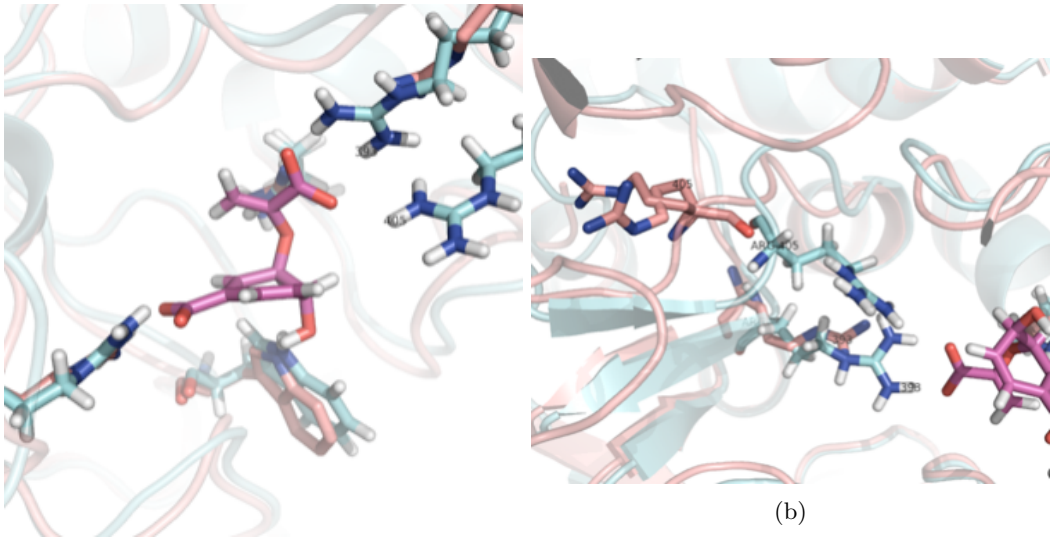



(a) (b)

Figure 4.5: ***In vitro* assays of Molecular Dynamics Hits (a)** SDS-PAGE gels of purified proteins used for assay **(b)** Activity assay for purified enzymes. The decrease in substrate concentration was monitored by looking at the absorbance at 275 nm.

The complementation assay uses a strain of *E. coli* that has its wild-type chorismate mutase activity knocked out. When grown in media without tyrosine and phenylalanine, cells that do not

have a plasmid encoding an active chorismate mutase enzyme are unable to grow. None of the designs was able to successfully complement the chorismate mutase activity of the *E. coli* strain used in this assay. On the other hand, cells that contained plasmid encoding the wild-type *E. coli* chorismate mutase and the wild-type sequence of one of the scaffolds used, 2GBB grew in the selective media. Wild-type 2GBB encodes another chorismate mutase from *Y. Pestis*. In order to account for the possibility that the *in vivo* complementation assay could be too stringent and that we had an active enzyme that was not efficient enough to complement activity in a knockout strain, we purified them for *in vitro* assays (Figure 4.5A). We used the assay conditions previously used by Lasilla *et. al* to measure the activity of different mutants of wild-type *E. coli* chorismate mutase [61]. We measured the decrease in absorbance at 275 nm for each of the designs (Figure 4.5B). Again, it was only 1ECM (wild-type *E. coli* chorismate mutase, and wild-type 2GBB that showed chorismate mutase activity.



(a) (b)

Figure 4.6: **X-ray Crystallography of Inactive Design. (a)** Comparison of active site of decoy (blue) and structure (red) one of the catalytic arginines is misplaced. **(b)** Interactions between R393 and R405 cause movement of the backbone displacing one of the catalytic residues.

In order to gain more information from the failed designs, we set up larger expression cultures of the best expressing mutants (1RQJ, 2B69, 2FVY_B, 2RJH, 3G5T, and 3QT9). Cells containing the proteins in pET-DEST expression vectors were grown in 1L cultures seeded by a 25 mL overnight starter culture at $37°C$ until they reached an optical density of 0.6. They were then induced with

100$\mu$M IPTG and moved to 22°$C$ and allowed to grow overnight ($\sim$ 16 hours). Cells were spun at 4000 RPM for 30 minutes. We resuspended the cells in buffer containing 150 mM NaCl, 50 mM, 10 mM imidazole, and 0.1 % Triton-X at pH 8.0. Cells were lysed by sonication (4:00 total ON time, with 3 second pulses and 5 second recovery). We spun the solution at 15000 RPM for 15 minutes and recovered the soluble protein with a nickel resin gravity column.

This was done to set up trays for x-ray crystallography of the designs. We looped around 30 crystals to send for x-ray crystallography. From these, we were able to collect a data set at 1.5 Å from one of the crystals corresponding to 3QT9.

Conditions for crystallization were as follows: the protein was concentrated at 13 mg/mL and the crystallant was 0.1 M CHES, and 30% PEG 3000 at a pH of 9.5. This was part of the Wizard I & II screen. We used parafin oil as our cryoprotectant.

When we compared the crystal structure to the design, we observed that the backbone had moved when compared to the starting structure we had used as a scaffold for our design, and this moved one of the catalytic arginines out of position from the design (Figure 4.6). This highlights one of the limitations of the design algorithm that was used which relies on a fixed backbone for rotamer placement. Based on the crystal structure, we decided to focus our enzyme design efforts on 3QT9– the protein expressed and crystallized well. Using the insights gained from the crystal structure, we ran a second round of computational protein design with 3QT9 as the starting scaffold.

### 4.3.6 Second Round of Designs based on 3QT9

Looking at the location of the active site for the initial 3QT9 enzyme design, we determined that the active site location may have been too solvent exposed. We selected a more buried arginine reside, R64, as the starting residue for Targeted Ligand Placement. We modified the geometry definitions for the energy biases that favored stabilizing interactions for the transition state in the design algorithm (Tables 4.3–4.6). Specifically, we changed the range of allowable distances and angles to be closer to the idealized molecular orbital geometries. We also allowed a greater set of interactions between amino acids and the transition state to be favored to include any potential

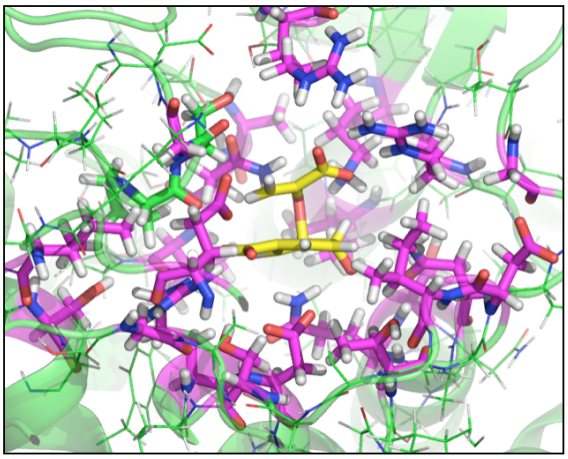


Figure 4.7: **One Redesign with the crystallized scaffold.** New designs were built off an arginine residue that was more buried than the initial designs. Interactions between residues were also looked at more carefully and geometry definitions were modified.

hydrogen bond interaction (Figure 4.7). We also used a higher resolution (0.9 Å) conformer library. While these changes increased the size of the search space for the computational design problem, we were able to use TRIAD, protein design software from Protabit to quickly explore this design space. We evaluated potential designs from the second round of calculations and chose 8 designs to run MD calculations on. Unfortunately, none of these designs were predicted to be active by the molecular dynamics prescreen.

Table 4.3: **Modified geometry definitions for first arginine interaction.** Geometry definitions for the bias function for the interaction between one of the catalytic arginines and CMX (the transition state of chorismate mutase reaction). [A, B] indicates the centroid of the coordinates of the two atoms specified. We allowed for the six possible interactions of the arginine with the substrate due to symmetry of the guanidino group: (A,B) = (NE,NH2), (NH1,NH2), or (NE, NH1). This was one of three required interaction for Targeted Ligand Placement.

| | CMX Atoms | Residue Atoms (Arginine) | Range |
|---|---|---|---|
| Distance | [O13,O14] | [A,B] | 2.50–3.30 Å |
| Angle 1 | [O13,O14] | [A,B], A | 70°–110° |
| Angle 2 | O13 or O14, [O13,O14] | [A,B] | 70°–110° |
| Dihedral 1 | [O13,O14] | [A,B], A, CZ | 160°–200° |
| Dihedral 2 | O13 or O14, [O13,O14] | [A, B], A | -20°–20° |
| Dihedral 3 | C11, O13 or O13, [O13,O14] | [A,B] | 160°–200° |

Table 4.4: **Modified geometry definitions for second arginine interaction.** Geometry definitions for the bias function for the interaction between one of the catalytic arginines and CMX (the transition state of chorismate mutase reaction). [A, B] indicates the centroid of the coordinates of the two atoms specified. We allowed for the six possible interactions of the arginine with the substrate due to symmetry of the guanidino group: (A,B) = (NE,NH2), (NH1,NH2), or (NE, NH1). This was one of three required interactions for Targeted Ligand Placement.

| | CMX Atoms | Residue Atoms (Arginine) | Range |
|---|---|---|---|
| Distance | [O11,O12] | [A,B] | 2.50–3.30 Å |
| Angle 1 | [O11,O12] | [A,B], A | 70°–110° |
| Angle 2 | O11 or O12, [O11,O12] | [A,B] | 70°–110° |
| Dihedral 1 | [O11,O12] | [A,B], A, CZ | 160°–200° |
| Dihedral 2 | O11 or O12, [O11,O12] | [A, B], A | -20°–20° |
| Dihedral 3 | C10, O11 or O12, [O11,O12] | [A,B] | 160°–200° |

Table 4.5: **Geometry definitions for hydrogen bond between protein and developing carbonyl of CMX.** Geometry definitions for the bias function for the interaction between a residue in scaffold and a developing carbonyl in the transition state. This oxygen is predicted to be negative so a lysine interaction is preferred. Atom identities for specific amino acids: lysine: A–NZ, glutamine: A–NE2, asparagine: A–ND2. This was one of the three required interactions for Targeted Ligand Placement.

| | CMX Atoms | Residue Atoms (Arginine) | Range |
|---|---|---|---|
| Distance | O7 | A | 2.50–3.30 Å |
| Angle 1 | C8 , O7 | A | 89.50°–129.50° |
| Angle 2 | C5 , O7 | A | 89.50°–129.50° |

Table 4.6: **Geometry definitions for hydrogen bond between protein and hydroxyl of CMX** Geometry definitions for the bias function for the interaction between a residue in scaffold and a hydroxyl in CMX. Atom identities for specific amino acids: serine: A–OG, B–HG, threonine: A–OG1, B–HG1, glutamine: A–NE2, B–1HE2 or 2HE2, asparagine: A–ND2, B–1HD2 or 2HD2

| | CMX Atoms | Residue Atoms (Arginine) | Range |
|---|---|---|---|
| Distance | O4 | A | 2.50–3.30 Å |
| Angle 1 | C4, O4 | A | 89.50°–129.50° |
| Angle 2 | O4 | B, A | 160°–180° |

## 4.4 Conclusion and Outlook

While the workflow failed to yield a successful design, this work contributed to the field by setting up a framework for a focused and systematic effort in computational enzyme design. One of the problems in enyzme design is the large space of potential proteins to assay. By applying multiple filters based on scientifically grounded hypotheses, we are able to drastically reduce to space of candidates to test into smaller problems that are experimentally tractable. The results of these experiments could then be looped back into the computational methods, providing new directions or new areas of the design space to explore (Figure 4.1). Advances in computational protein design, such as design with moveable backbones, coupled with advances computing power that allow us to explore more of the design space, will increase the efficacy of this workflow by providing us with more accurate answers towards which parts of the space to explore experimentally.

# Chapter 5

# Conclusion and Future Directions

## 5.1 Summary of Contributions

Synthetic biology has the potential to be a transformative technology. Particularly exciting is the impact it can have in the discovery and production of pharmaceuticals. The research required in order for this to occur draws from many different disciplines. Part of the work required is the development of frameworks that allow synthetic circuits to behave more like natural circuits. This thesis lays some of groundwork by proposing and implementing new mechanisms in which this can occur and are steps towards this goal.

In Chapter 2, we proposed and implemented a mechanism in which cells can talk to a synthetic circuit by engineering a transcription factor to sense vanillin, a small molecule that is relevant in biofuel fermentation. We see this process as a framework that can be implemented to create transcription factors that are able to transmit information in other synthetic circuits leading to dynamic regulation of a synthetic circuit by molecules specifically tied to that circuit, mimicking the behavior of natural circuits. To the best of our knowledge, this is the first instance of rational transcription factor engineering where the target effector was unrelated to the ligands that the native transcription factor was sensitive to. QacR and vanillin were chosen independently and the framework is agnostic to the choice of small molecule and transcription factor to engineer.

Chapter 3 demonstrates a mechanism for dynamically fixing the ratio between two proteins. Mathematical modeling suggests that this ratio is tunable by altering the levels of response regulator

and phosphatase, which change the gain of the system and are independent of the proteins whose ratios we wish to control. This addresses the problem that arises when proteins from different organisms are placed together in a synthetic circuit. Since they do not operate together in a natural context, the output of one enzyme may not be optimized to the input of the succeeding enzyme, and fixing the ratios between these proteins allows their productivity to match more closely and can increase yields in a synthetic metabolic network [38]. The work here expands this by combining a feedback loop with scaffold-dependent phosphorylation and we allow the proteins to interact and maintain their ratio with changing environmental conditions. To the best of our knowledge, this is the first time that dynamic tracking of a signal was achieved entirely in the context of the cell, since the components of the tracker were proteins.

While unsuccessful in developing an enzyme that had novel chorismate mutase activity, Chapter 4 lays a framework for tackling enzyme engineering challenges in a systematic way. This framework is flexible and allows for advances in both computational resources and advances in computational protein design. We can readily apply this framework for other reactions that have a known mechanism and target. Reactions must be chosen so they are chemically feasible, but do not have to be naturally observable. With the proper choice of target reactions, we can expand the space that is chemically accessible to synthetic metabolic networks, further increasing the potential of synthetic biology.

## 5.2 Future Directions

One can imagine combining the different components in this thesis to create a synthetic metabolic network that performs like a natural gene network. This metabolic network would respond to changing conditions regulating genes based on the amount needed to process a substrate (Chapter 2). Enzymes in the network would maintain the optimal ratio for maximum target yield (Chapter 3) and the target could be a molecule that is not naturally produced by any organism (Chapter 4). The ability to do this rationally and systematically for many targets would mean synthetic biology fully realizing its potential. Much work is necessary in order to achieve this.

Applying the framework for switching effector specificity on another target small molecule or another transcription factor would increase our confidence that we can use this as a general method to communicate with our synthetic circuits. Using the vanillin sensor specifically in a closed feedback loop to allow the cell to respond to vanillin toxicity would also show that dynamical responses are beneficial to cells; an engineered stress response to vanillin is something we are interested in pursuing.

More characterization of the biomolecular concentration tracker to understand the amount of control we have on the ratio between the two proteins will be necessary in order to use it as a mechanism for control in a synthetic metabolic circuit. The addition of another feedback loop may also be necessary to allow the system to respond to decreases in small molecule concentration, as well as increases, so as not to overburden the cell's protein degradation machinery. The mechanism can also be extended to maintain the ratio of more than two proteins, since metabolic circuits are rarely composed of only two enzymes.

The framework for computational enzyme design has to be successfully applied to the design of an enzyme in a metabolically relevant pathway. It can also be modified to take advantage of the more recent advances in computational enzyme design.

Finally, it would be advantageous to integrate more modeling of the different components that compose a synthetic metabolic pathway. This can be a combination of metabolic flux models and differential equation based models of protein-protein, protein-DNA, and protein-small molecule interactions. Oftentimes, optimization and design of synthetic circuits is done by screening large libraries of combinations promoters and ribosomal binding sites, essentially randomly searching design space to find the combination that gives the desired behavior. Integrating more systems based modeling could help address this.

While there is a lot of work still to be done, the research described here is hopefully a step in the right direction towards fully realizing the benefits we can attain from synthetic biology.

# Appendix A

# Engineering Post-Translational Modification Switchable Domains for Fast, Programmable Allosteric Feedback

## A.1 Introduction

Regulators such as pBAD [62] and pLac [63] are an important part of the synthetic toolkit because they allow for control of output in synthetic circuits. However, both these regulators operate on the DNA-level and modulate their downstream signals by allowing the production of new proteins. This mechanism causes responses from inputs to occur at the slower time scales. In natural systems, phosphorylation and other post-translational modifications are used when a rapid response to a signal is necessary. A classic example of this is the chemotaxis circuit in *E. coli*, which is governed by a series of protein signaling cascades mediated by protein-protein interactions and post-translational modifications, namely, phosphorylation and methylation [64]. By operating at the protein level, *E. coli* is able to rapidly respond to an external signal (a chemoattractant) in the seconds time scale. To be able to achieve similar performance, synthetic circuits must incorporate control elements at the protein level and use "protein switches."

The insertion of ligand-binding domains into different types of proteins allows one to tune a proteins function using triggers such as calcium [65], maltose [66], or small peptides [67]. Erster

and colleagues showed that inserting a tetracysteine (4C) motif into different proteins results in a FlAsH-dependent modulation of activity: catalytic activity was altered by about 5-fold in TEM-1 $\beta$-lactamase [68] and seven-fold in ERK [69], a yeast MAP Kinase. Reynolds and colleagues have also developed a light sensitive dihydrofolate reductase enzyme (DHFR) by inserting a photo-sensitive LOV domain into DHFR [70]. These results suggest that domain insertions can provide a general method to modulate protein activity. Although in most cases the exact mechanism of activity modulation is unknown, it is possible that structural changes in the inserted domain due to ligand binding propagate throughout the protein altering its activity.

While ligand-sensitive protein switches operate on the protein level, their generalizability is limited because of their dependence on a specific small molecule. Furthermore, the use of small molecules makes reversibility of the switch problematic. In natural systems such as the two-component systems in bacteria and the MAPK cascades in eukaryotes, external signals (including small molecules) are processed and converted into signaling cascades that use post-translational modifications such as phosphorylation to transmit the signal through a protein network.

Post-translational modifications effect allosteric changes on their target proteins, thereby altering their functionality. Protein phosphorylation causes conformational changes both in natural [71] and synthetic systems [72]. Riemen and Waters designed the Trp-switch, a short peptide that transitioned from a $\beta$-sheet to a random coil upon phosphorylation of a serine residue. This result showed that designing short phospho-sensitive peptides that are enzymatic targets is feasible. Signarvic and DeGrado used computationally aided protein design with an idealized coiled coil model to design sequences that tetramerize upon phosphorylation of an N-capping serine residue [73].

This chapter describes my efforts to engineer a general mechanism for allostery in proteins. The way I proposed to do this was by engineering a phosphorylatable domain that had two different conformations in its phosphorylated and non-phosphorylated states. The hypothesis was that inserting this domain into an enzyme would result in phosphorylation-dependent modulation in activity, as structural changes in the domain would propagate through the enzyme and cause changes in its activity (Figure A.1).

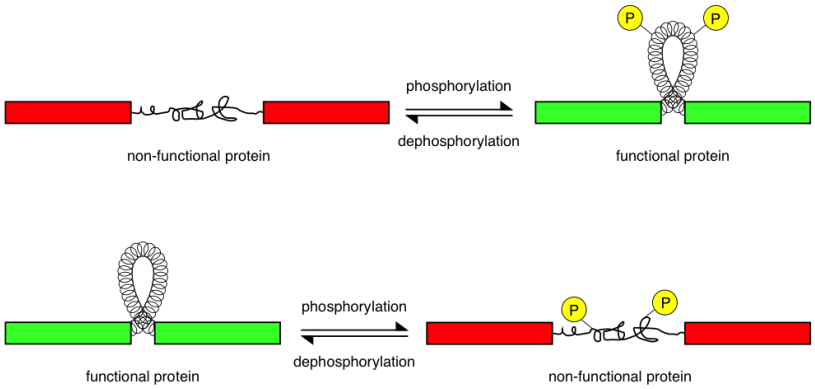



Figure A.1: **Schematic representation of programmable protein structure.** A linker is inserted between two sections of a naturally occurring protein. In the top configuration, phosphorylation of the linker will bring together the sections of the original protein such that the protein is functional. In the bottom configuration, phosphorylation has the opposite effect.

## A.2 Characterizing and Finding Potential Enzymes for Activity Assays

Wanting to build on previous work, I looked at enzymes that had either been sucessfully split and reconstituted, or had been active after domain insertions. I first attempted to insert the Trp-Switch [72], a 12 amino acid long peptide designed by Riemen and Waters [72], between amino acids 96 and 97 (TEM-1 96) and amino acids 215 and 216 (TEM-1 215) of *E. coli* $\beta$-lactamase. Erster *et al.* had previously shown that insertion of a tetracysteine domain into these locations caused a FlAsH-dependent modulation in activity [68]. Rieman and Waters demonstrated that the Trp-switch, normally in a $\beta$-hairpin conformation, becomes unstructured upon phosphorylation of a serine residue by protein kinase A (PKA). By inserting this short peptide into locations that have previously been demonstrated to be amenable to peptide insertions that could induce changes in enzyme activity, I hoped to demonstrate some modulation in $\beta$-lactamase activity due to the phosphorylation of the Trp-switch. I was unable to successfully purify the protein with the Trp-switch insertions for further characterization. Expression at both 18°C and 37°C at IPTG induction levels of 1 mM and 0.01 mM did not yield soluble protein. I attempted to refold the protein from the insoluble fraction using the protocol described by Erster *et al.* [68]. However, while I was able

to successfully refold the wild-type protein, the protocol did not successfully refold the proteins containing the domain insertions. One of the requirements I wanted for the enzyme I used in the assay was its compatibility with medium and high throughput screens. A refolding step before the assay would significantly reduce the throughput of any screen. Because of this, I looked at other proteins.

The second protein I explored was luciferase from *P. pyralis*. Luciferase was an interesting candidate because of the modularity of its N and C-terminal domains demonstrated by the use of split luciferase in the study of protein-protein interactions. I cloned genes encoding the wild-type firefly luciferase and a Trp-switch domain inserted luciferase into expression vectors. The domain inserted luciferase construct consisted of an N-terminal domain (residues 1-416), followed by the Trp-switch, and the C-terminal domain (residues 398-550). The luciferase fragments I selected had previously been shown to give high signal-to-noise ratios in an FRB/FKBP protein interaction study [74]. I was able to successfully purify both constructs. Both constructs were shown to be folded (by circular dichroism) and active (by a luciferin activity assay). However, when I tried to set up reaction conditions to phosphorylate the domain inserted construct, it crashed out of solution. The general instability of the luciferase constructs led me to look for another model enzyme to assay.

I finally settled on using *E. coli* dihydrofolate reductase (DHFR). DHFR had previously been made light-sensitive by Reynolds and colleagues by inserting photo-responsive LOV domains in three different locations [70]. Using DHFR provided many advantages over the previous enzymes tried due to its small size ($\sim$ 20 kDa), simple purification protocol, and overall general stability. DHFR activity could be assayed by an *in vitro* plate assay that measured the decrease in absorbance at 340 nm, which tracks the reduction of NADPH to NADH, a co-factor in the reaction, or an *in vivo* cell growth assay [70].

## A.3 Designing a Phosphoswitchable Domain

Since we cannot directly design for the phospho-switching activity, I focused on domains with target structures that would potentially correlate to the function we wished to design. In this case, I

wanted to maximize the energy gap between a folded state with a short end-to-end distance and an unfolded state with a longer end-to-end distance with phosphorylation as the trigger between these two states. The underlying hypothesis is that when in the folded (structured) state, the engineered domain would have a short end-to-end distance, allowing for its non-deleterious insertion into a protein. We want to strategically select the domain insertion points such that they are locations at the surface of the protein that are potentially sensitive such that unfolding of the insertion domain will propagate through and destabilize the entire protein. Reynolds *et al.* demonstrated the existence of these insertion points by using co-evolutionary analysis of different surface residues in DHFR to find "hotspot" residues that were on the surface of DHFR but were connected to a network of residues linking to the active site of the protein [70]. Another design constraint that I included was that the target residues for phosphorylation must be targets for enzymatic phosphorylation by PKA. Enzymatic phosphorylation of domains is required to increase throughput of the screening. PKA was selected because it is a relatively well-characterized kinase, its consensus phosphorylation site is relatively flexible (R-R-X-S/T-Φ) [75], and because its active form can be expressed in E. coli [76]. The

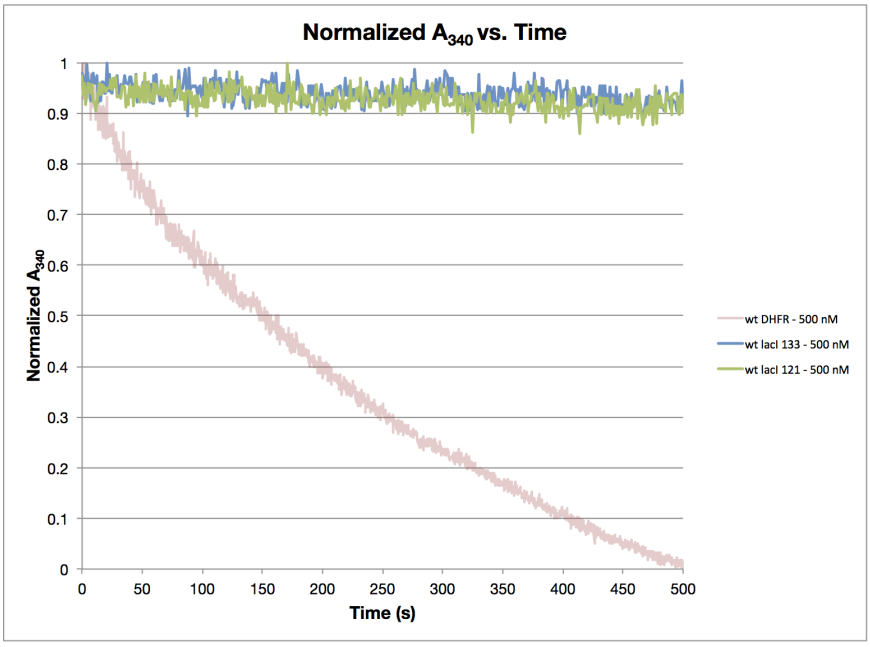



Figure A.2: **No detectable DHFR activity with the LacI Tetramerization Domain Inserted** Absorbance at 340 nm of wild-type DHFR, or DHFR with an insertion of a single-chain version of the LacI tetramerization domain. A functional enzyme will correspond to a decrease in absorbance as NADPH is reduced as part of the reaction. The domain insertions have no detectable activity.

first domain I tried was based on the tetramerization domain of the lacI protein (LacI-tet). When lacI is a monomer, this domain is mostly unstructured; however, when the protein tetramerizes, the monomers adopt an anti-parallel four-helix bundle conformation. Signarvic and DeGrado created variants of LacI-tet that promoted tetramerization upon enzymatic phosphorylation of an N-capping serine residue; phosphorylation serves to stabilize two arginines by balancing their charges, allowing the monomers to become more structured and tetramerize [73]. The phosphorylated peptides they designed favored tetramerization with an energy difference of up to $-4.6$ kcal/mol when compared to their unphosphorylated state. To see if I could use a similar concept as a starting point, I designed single-chain versions of the wild-type lacI tetramerization domain (sc-wtLacI-tet) and the mutant tetramerization domain that was phosphorylation sensitive (sc-RRSLacI-tet) and inserted them into DHFR. The insertion of these domains into DHFR and sites determined by the Reynolds [70] caused the enzyme to lose any measurable activity (Figure A.2). Circular Dichroism conducted on the purified sc-wtLacI-tet domain as a separate peptide suggested that it was stable and folded. I was unable to express sc-RRSLacI-tet on its own. I hypothesized that the domain insertion was too big for the enzyme to accomodate and started to look for alternative domains. I wanted to continue exploring

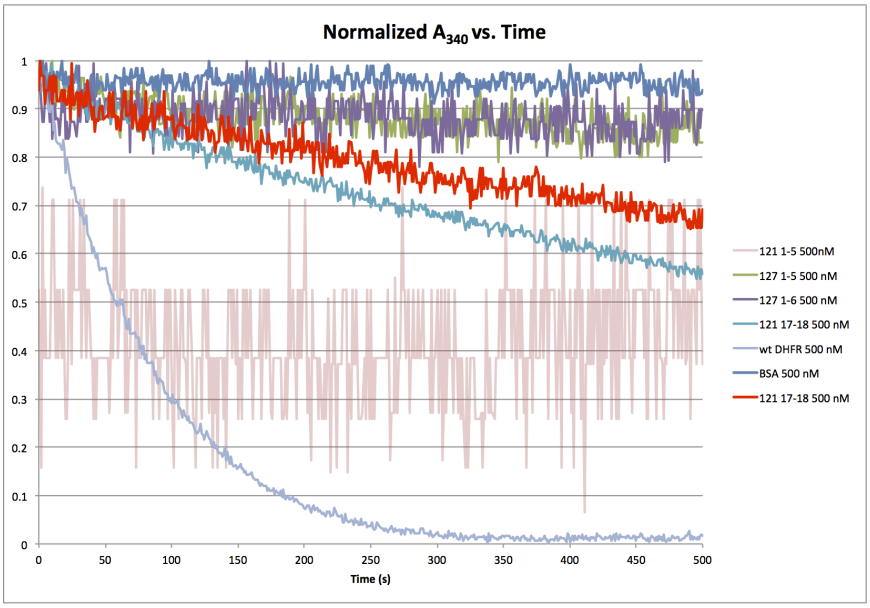



Figure A.3: **DHFR with Interacting Leucine Zipper Domain Insertion Shows Activity.** Activity was detected in DHFR containing a domain insertion with interacting leucine zippers at two different insertion points (red and cyan). No activity was detected when non-interacting leucine zippers were inserted.

the potential of using alpha-helical interactions as the mechanism for phosphoswitchable activity. The Keating lab had characterized a set of synthetic leucine zipper interactions, and they specifically looked at the different properties in their coiled-coil interactions, including cross-talk and directionality (parallel vs. anti-parallel) [77]. I focused on the Synzip 17-Synzip 18 interaction (17+18), since it was the only anti-parallel interaction characterized. This made it amenable for insertion into an enzyme as a single chain coiled-coil domain. To see if I could use leucine zipper dimerization as a mechanism for creating phosphoswitchable domains, I inserted single-chain versions of leucine zippers that I obtained from the Keating lab that they had determined to be interacting (17+18) or non-interacting (1+5 and 1+6) into DHFR. If leucine zipper dimerization was a mechanism that warranted further exploration, we would expect to see activity when 17+18 was inserted into DHFR, and observe less or no enzyme activity when 1+5 or 1+6 was inserted. This is what was observed. When I inserted the domains in the locations of DHFR previously determined by Reynolds *et al.*, I detected DHFR activity for the enzymes with 17+18 inserted and no activity for the enzymes with 1+5 or 1+6 inserted (Figure A.3). This was an encouraging result and I started to look more closely at the 17+18 interaction to find ways to stabilize and destabilize it using phosphorylation.

## A.4 Modifying the 17+18 Domain to be Sensitive to Phosphorylation

Having demonstrated the potential of having the interaction between leucine zipper domains mediate changes in enzyme activity. I began to examine potential mutations to synzips 17 and 18 to both introduce the PKA recognition motif (RRx**S/T**Φ) and provide a change in stability upon phosphorylation of the key residue in the motif. I put the PKA phosphorylation domain at the N-terminus of each of the leucine zippers. This is the strategy used by Signarvic and deGrado to achieve a difference in stability for tetramerization of the lacI-derived monomers between the phosphorylated and non-phosphorylated states [73]. The PKA phosphorylation domain contains positively charged arginines. Placing it on the N-terminus of an alpha-helix increases the dipole moment and destabi-

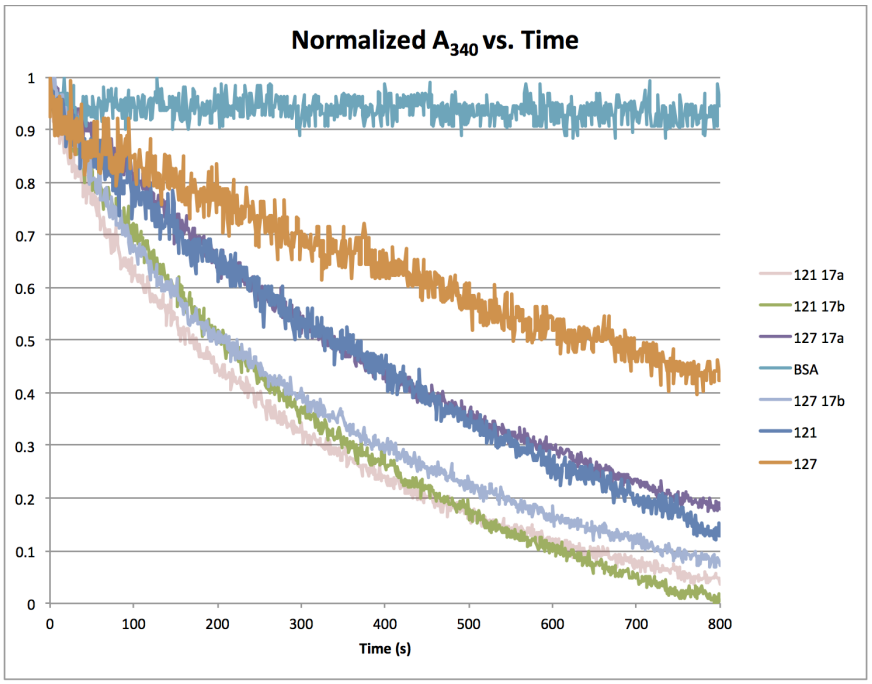



Figure A.4: **Mutations to 17+18 sequences do not decrease enzyme activity.** DHFR activity was measured for enzymes inserted with modified 17+18 sequences (17a and 17b) at two different locations (121 and 127).

lizes the helix. Phosphorylation of the serine residue in the motif adds negative charge to stabilize the helix. Further stability is added by hydrogen bonding between the phosphate and the guanadinium groups of the arginine. In this case, we expect the enzymes that have the domains inserted to be less active when non-phosphorylated as phosphorylation of the serine residues in the N-terminus of the leucine zippers would stabilize the coiled coil interaction. However, when I assayed the enzymes containing the the modified 17+18 inserts, the activity was similar or higher than enzymes that contained the original 17+18 sequences (Figure A.4). We hypothesized that the coiled-coil might have been too stable since the measured melting temperature of the peptide was greater than 95 °C. The original 17+18 synzip from the Keating lab was six heptads long and consisted of forty-two interacting residues. I cut out two heptads from each of the leucine zippers to try to destabilize the coiled-coil interaction. I inserted the truncated 17+18 peptide (17+18s) into DHFR. I was only able to detect DHFR activity for one of the insert locations (before the 121st residue, DHFR 121). After confirming the activity of the 17:18s DHFR 121 protein, I proceeded to mutate residues of at the N-terminii of the leucine zippers to introduce a PKA recognition motif and destabilize the

helix. I inserted two different versions of the domain into DHFR, one in which the second residue of each of the leucine zippers was a serine (this was the target residue for phosphorylation by PKA (DHFR 121 RRS)); the second version had glutamate as the second residue of the leucine zippers (DHFR 121 RRE). Glutamate is commonly used as a mimic for a phosphorylated serine. Differences in activity between RRS and RRE would suggest that it was possible to get modulation in activity by phosphorylation of the serine residue in RRS. I was able to observe some difference in activity between RRS and RRE (Figure A.5). RRS appeared to be less active than RRE, consistent with our hypothesis that additional positive charge on the N-terminus of the helix would increase the dipole and lead to instability of the coiled-coil interaction. However, when I set up a reaction where I first incubated the enzyme with PKA, I did not see any difference in activity between RRS that was pre-incubated in PKA or blank reaction buffer (Figure A.6). I hypothesized that this was because the enzyme was not getting phosphorylated. I had difficulty in testing this hypothesis.

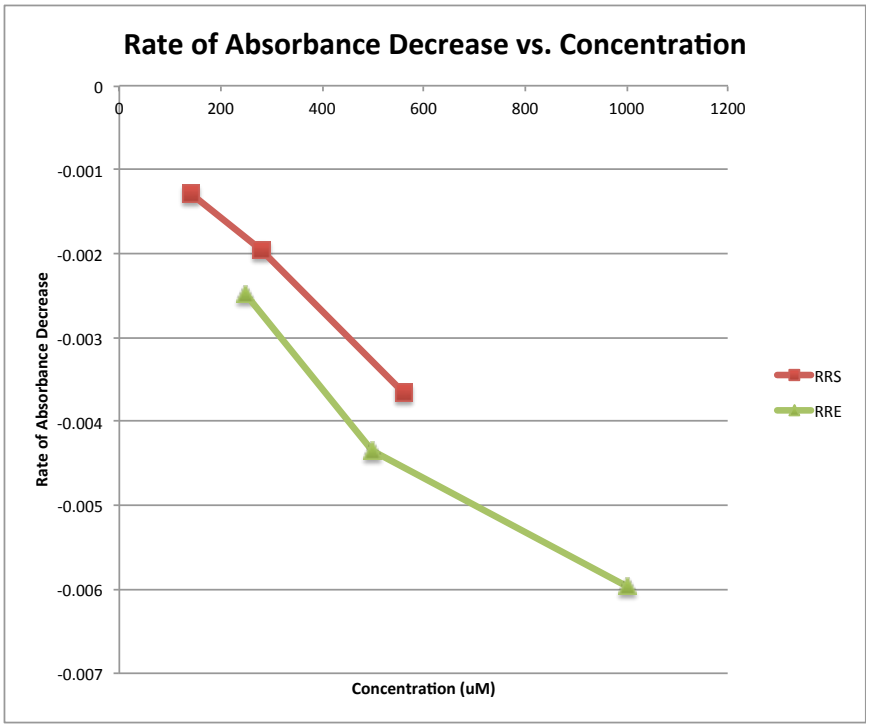



Figure A.5: **Apparent differences in activity between DHFR 121 RRS and DHFR 121 RRE.** The rate in decrease of absorbance was plotted for different concentrations of DHFR with domain insertions with either an unphosphorylated (RRS) or phospho-mimic (RRE) variant of a leucine zipper pair. The enzyme with the RRS insert appears to be less active, consistent with our hypothesis that positive charges on the N-terminus of the leucine zipper will destabilize the domain.

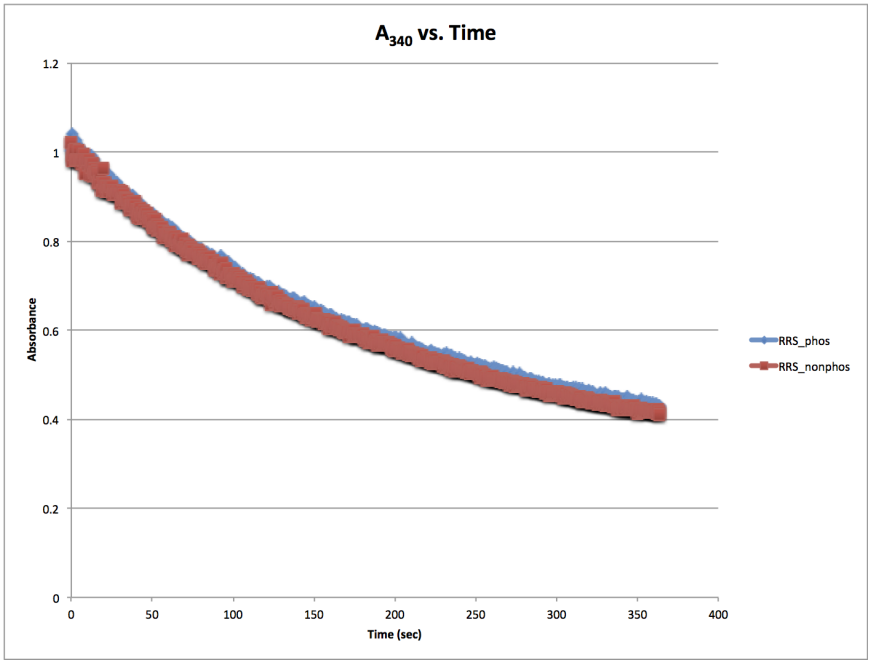



Figure A.6: **No Difference in Activity Between DHFR 121 RRS incubated with and without PKA.** Purified DHFR 121 RRS was preincubated with PKA or blank reaction buffer. There was no difference observed in DHFR activity.

## A.5 Difficulties in Phosphorylating Enzyme and Assaying for Phosphorylation Made Progress Difficult

I could not find a good way to test for PKA activity. The idealized substrate of PKA, the Kemptide, was a small 7 amino acid peptide that was 772 daltons. It was difficult to trap in an acrylamide gel when I tried a radio-labeled phospho-transfer assay. I tried liquid-chromatography mass spectrometry, but I could not find a column that would trap the phosphorylated version of the kemptide. The size of the peptide and my uncertainty about the antibody I was using also made it difficult to find the right conditions for Western Blotting.

Another alternative I explored was to place the phosphoserine in ribosomally using a system developed in the Rinehart lab that consisted of a phosphoserine amber tRNA–tRNA synthetase pair [78]. This would allow me to assay differences in enzyme activity between the enzyme with and without phosphorylated serine residues, independent of enzymatic phosphorylation. In my hands, I was not able to detect protein expression of any of the DHFR mutants, or a sfGFP amber

mutant via commassie staining using either a BL21ΔserB *E. coli* strain, or an EcAR7.SEP strain, which has RF-1 deleted and some amber and opal mutations that showed an increase in phosphoserine insertion in purified proteins [79]. Discussions with the Rinehart lab confirmed that overall expression even with the modified strain were low and that Western Blots were typically necessary to detect protein. I ultimately decided to abandon this approach because the length of time required to express the purified proteins, coupled with the low yields, would make obtaining enough protein for assays difficult. Furthermore, even if we were able to detect differences in enzyme activity using this approach the designs would have to be modified to be compatibile for enzymatic phosphorylation regardless for use in a synthetic circuit. While I was working on a method to detect phosphorylation

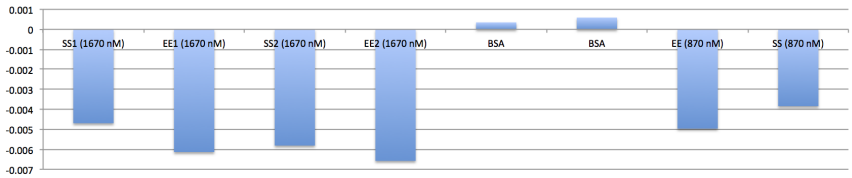



Figure A.7: **Activity assay for modified DHFR 121 enzymes with idealized phosphorylation sites.** Rate of decrease of absorbance at 340 nm for the first 50 seconds of the reaction for two different enzyme concentrations. DHFR with the phospho-mimic domain inserted (EE) appears to be more active than DHFR with the unphosphorylated domain inserted (SS).

of residues, I also looked at ways that I could optimize the kinase recognition sequences to increase the likelihood that the enzymes with domain inserts could be enzymatically phosphorylated. Songyang *et. al* used a peptide library to determine the optimal substrate for PKA [80]. Looking at their data, I decided to mutate the preceding residue of the recognition domain to arginine, and the residue after the target serine to an isoleucine. When I compared unphosphorylated versions of these enzymes with their corresponding phospho-mimics, I observed the difference in activity in the modified PKA motifs (Figure A.7).

## A.6 Potential Next Steps

In this section, I briefly outline some of the potential next steps that I believe are worth exploring to continue research in this project.

### A.6.1 *In Silico* Pre-screen Using Molecular Dynamics

Dagliyan and colleagues used Replica Exchange Discrete Molecular Dynamics (RexDMD) to design a domain based on the rapamycin-dependent interaction between FKB12 (iFKBP)–FKBP12 and rapamycin binding protein (FRB) [81]. Starting with iFKBP, which had previously been shown to maintain its structure when inserted as a domain, they added the rapamycin-binding surface of FRB. They also grafted helices from FRB to add stability to the domain. They used RexDMD to computationally test the stability of their domain in the presence and absence of rapamycin. Specifically, they used RexDMD to estimate the termini distance and specific heat of the domain. By looking at the shift in the peak of the specific heat curve, which would correspond to a folding transition in the presence and absence of rapamycin, they were able to evaluate whether their domain had the desired properties computationally before experimental testing. One could use a similar approach and simulate the coiled-coil interaction between phosphorylated and non-phosphorylated versions of the domain.

### A.6.2 High-throughput Screening Assay for DHFR activity

Instead of measuring the absorbance at 340 nm, which is time sensitive and requires monitoring immediately after injection of substrate, Kumar and his colleagues developed a high throughput coupled assay for DHFR activity using diaphorase and resazurin [82]. The DHFR reaction is run for a set time and then stopped. Diaphorase consumes the remaining NADPH to make a colored product, the emission of the product is measured and is used as a proxy for DHFR activity. Setting up reactions in this way would enable one to test more candidate designs. However, the bottleneck for needing to purify the enzyme and normalize the concentration for the reaction is still present.

## A.7 Outlook for the Project

The project was hampered by the difficulty in testing for enzymatic phosphorylation. I have no clear solution to this problem. I have not tried dot blotting with the anti-phosphoserine antibody as

a method to detect the phosphorylated peptide. If that is unable to detect the difference between the kemptide and phosphorylated kemptide, it is not clear what the next step would be. The lack of a good assay led me to abandon this project. Furthermore, even if this project were successful, there still exists the problem of fast signaling. One would need a signal that operates at a similarly fast time scale, that would turn on the kinase to modulate the protein activity. Otherwise, the bottleneck for a slow response would still exist. Recent advances in tunable protein degradation [83] and light-based activation and sequestration [70, 84] from when I started the project seem to be more promising avenues to get modulation of activity of enzymes at a fast time scale.

# Bibliography

[1] DK Ro, EM Paradise, M Ouellet, KJ Fisher, KL Newman, JM Ndungu, KA Ho, RA Eachus, TS Ham, J Kirby, MCY Chang, ST Withers, Y Shiba, R Sarpong, and JD Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, April 2006.

[2] J Turconi, F Griolet, R Guevel, G Oddon, R Villa, A Geatti, M Hvala, K Rossen, R Göller, and A Burgard. Semisynthetic artemisinin, the chemical path to industrial production. *Organic Process Research & Development*, 18(3):417–422, 2014.

[3] VJJ Martin, DJ Pitera, ST Withers, JD Newman, and JD Keasling. Engineering a mevalonate pathway in escherichia coli for production of terpenoids. *Nature Biotechnology*, 21(7):796–802, July 2003.

[4] CJ Paddon, PJ Westfall, DJ Pitera, K Benjamin, K Fisher, D McPhee, MD Leavell, A Tai, A Main, D Eng, DR Polichuk, KH Teoh, DW Reed, T Treynor, J Lenihan, H Jiang, M Fleck, S Bajad, G Dang, D Dengrove, D Diola, G Dorin, KW Ellens, S Fickes, J Galazzo, SP Gaucher, T Geistlinger, R Henry, M Hepp, T Horning, T Iqbal, L Kizer, B Lieu, D Melis, N Moss, R Regentin, S Secrest, H Tsuruta, R Vazquez, LF Westblade, L Xu, M Yu, Y Zhang, L Zhao, J Lievense, PS Covello, JD Keasling, KK Reiling, NS Renninger, and JD Newman. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446):528–532, April 2013.

[5] C Marris. Synthetic biology's malaria promises could backfire, October 2013. Online

URL:http://www.scidev.net/global/biotechnology/opinion/synthetic-biology-s-malaria-promises-could-backfire.html. Accessed April 21 2015.

[6] JR Doroghazi, JC Albright, AW Goering, KS Ju, RR Haines, KA Tchalukov, DP Labeda, NL Kelleher, and WW Metcalf. A roadmap for natural product discovery based on large-scale genomics and metabolomics. *Nature Chemical Biology*, 10(11):963–968, November 2014.

[7] CW Johnston and NA Magarvey. Natural products: Untwisting the antibiotic'ome. *Nature Chemical Biology*, 11(3):177–178, March 2015.

[8] AS Khalil and JJ Collins. Synthetic biology: applications come of age. *Nature Reviews Genetics*, 11(5):367–379, May 2010.

[9] F Zhang, JM Carothers, and JD Keasling. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology*, 30(4):354–359, April 2012.

[10] PEM Purnick and R Weiss. The second wave of synthetic biology: from modules to systems. *Nature Reviews. Molecular Cell Biology*, 10(6):410–422, June 2009.

[11] JT Meyerowitz. Engineering synthetic microbial stress responses. Unpublished Candidacy Report, 2011.

[12] TA Desai, DA Rodionov, MS Gelfand, EJ Alm, and CV Rao. Engineering transcription factors with novel dna-binding specificity using comparative genomics. *Nucleic Acids Research*, 37(8):2493–2503, 05 2009.

[13] TC Galvo and V de Lorenzo. Transcriptional regulators la carte: engineering new effector specificities in bacterial regulatory proteins. *Current Opinion in Biotechnology*, 17(1):34 – 42, 2006. Analytical biotechnology Edited by Jan Roelof van der Meer and J Colin Murrel.

[14] ELC de los Santos, JT Meyerowitz, SL Mayo, and RM Murray. Engineering transcriptional regulator effector specificity using computational design and in vitro rapid prototyping: Developing a vanillin sensor. *bioRxiv*, 2015. DOI: 10.1101/015438.

[15] JL Ramos, M Martinez-Bueno, AJ Molina-Henares, W Terán, K Watanabe, XD Zhang, MT Gallegos, R Brennan, and R Tobes. The TetR family of transcriptional repressors. *Microbiology and Molecular Biology Reviews*, 69(2):326–356, June 2005.

[16] MB Elowitz and S Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, 2000.

[17] TS Gardner, CR Cantor, and JJ Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

[18] S Grkovic, KM Hardie, MH Brown, and RA Skurray. Interactions of the QacR multidrug-binding protein with structurally diverse ligands: implications for the evolution of the binding pocket. *Biochemistry*, 42(51):15226–15236, 2003.

[19] MA Schumacher, MC Miller, S Grkovic, MH Brown, RA Skurray, and RG Brennan. Structural mechanisms of QacR induction and multidrug recognition. *Science*, 294(5549):2158–2163, December 2001.

[20] MA Schumacher and RG Brennan. Deciphering the molecular basis of multidrug recognition: Crystal structures of the Staphylococcus aureus multidrug binding transcription regulator QacR. *Research in Microbiology*, 154(2):69–77, March 2003.

[21] KM Peters, BE Brooks, MA Schumacher, RA Skurray, RG Brennan, and MH Brown. A single acidic residue can guide binding site selection but does not govern QacR cationic-drug affinity. *Public Library of Science ONE*, 6(1):e15974, 2011.

[22] HB Klinke, Thomsen, AB, and BK Ahring. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Applied Microbiology and Biotechnology*, 66(1):10–26, November 2004.

[23] JK Lassila, HK Privett, BD Allen, and SL Mayo. Combinatorial methods for small-molecule placement in computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45):16710–16715, November 2006.

[24] BD Allen and SL Mayo. An efficient algorithm for multistate protein design based on FASTER. *Journal of Computational Chemistry*, 31(5):904–916, April 2010.

[25] ZZ Sun, CA Hayes, J Shin, F Caschera, RM Murray, and V Noireaux. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of Visualized Experiments : JoVE*, (79):e50762, 2013.

[26] ZZ Sun, E Yeung, CA Hayes, V Noireaux, and RM Murray. Linear dna for rapid prototyping of synthetic biological circuits in an escherichia coli based tx-tl cell-free system. *ACS Synthetic Biology*, 3(6):387–397, 2014.

[27] BC Stanton, V Siciliano, A Ghodasara, L Wroblewska, K Clancy, AC Trefzer, JD Chesnut, R Weiss, and CA Voigt. Systematic Transfer of Prokaryotic Sensors and Circuits to Mammalian Cells. 3(12):880–891, 2014. PMID: 25360681.

[28] TN Nguyen, QG Phan, LP Duong, KP Bertrand, and RE Lenski. Effects of carriage and expression of the tn10 tetracycline-resistance operon on the fitness of escherichia coli k12. *Molecular Biology and Evolution*, 6(3):213–225, 1989.

[29] AD Bochevarov, E Harder, TF Hughes, JR Greenwood, DA Braden, DM Philipp, D Rinaldo, MD Halls, J Zhang, and RA Friesner. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *International Journal of Quantum Chemistry*, 113(18):2110–2142, July 2013.

[30] B Kuhlman, G Dantas, GC Ireton, G Varani, BL Stoddard, and D Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–1368, November 2003.

[31] BD Allen and SL Mayo. Dramatic performance enhancements for the FASTER optimization algorithm. *Journal of Computational Chemistry*, 27(10):1071–1075, July 2006.

[32] R Lutz and H Bujard. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25(6):1203–1210, 1997.

[33] E Franco, PO Forsberg, and RM Murray. Design, modeling and synthesis of an in vitro transcription rate regulatory circuit. In *American Control Conference (ACC)*, pages 2786–2791, 2008.

[34] ELC de los Santos, V Hsiao, and RM Murray. Design and implementation of a biomolecular circuit for tracking protein concentration. In *American Control Conference (ACC)*, pages 2290–2294, 2013.

[35] V Hsiao, ELC de los Santos, WR Whitaker, JE Dueber, and RM Murray. Design and implementation of a biomolecular concentration tracker. *ACS Synthetic Biology*, 4(2):150–161, 2015. PMID: 24847683.

[36] J Ang, S Bagh, BP Ingalls, and DR McMillen. Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network. *Journal of Theoretical Biology*, 266(4):723–738, October 2010.

[37] JA Stapleton, K Endo, Y Fujita, K Hayashi, M Takinoue, H Saito, and T Inoue. Feedback control of protein expression in mammalian cells by tunable synthetic translational inhibition. *ACS Synthetic Biology*, 1(3):83–88, March 2012.

[38] TS Moon, JE Dueber, E Shiue, and KLJ Prather. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metabolic Engineering*, 12(3):298–305, 2010.

[39] WR Whitaker, SA Davis, AP Arkin, and JE Dueber. Engineering robust control of two-component system phosphotransfer using modular scaffolds. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):18090–18095, October 2012.

[40] ES Groban, EJ Clarke, HM Salis, SM Miller, and CA Voigt. Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of Molecular Biology*, 390(3):380–393, 2009.

[41] WR Whitaker and JE Dueber. Metabolic pathway flux enhancement by synthetic protein scaffolding. *Methods in Enzymology*, 497:447–468, 2011.

[42] U Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits (Chapman & Hall/CRC Mathematical and Computational Biology)*. Chapman and Hall/CRC, 1 edition, July 2006.

[43] Y Pazy, AC Wollish, SA Thomas, PeJ Miller, EJ Collins, RB Bourret, and RE Silversmith. Matching biochemical reaction kinetics to the timescales of life: structural determinants that influence the autodephosphorylation rate of response regulator proteins. *Journal of Molecular Biology*, 392(5):1205–1220, October 2009.

[44] E Solomaha, FL Szeto, MA Yousef, and HC Palfrey. Kinetics of Src homology 3 domain association with the proline-rich domain of dynamins: specificity, occlusion, and the effects of phosphorylation. *The Journal of Biological Chemistry*, 280(24):23147–23156, June 2005.

[45] WR Whitaker, SA Davis, AP Arkin, and JE Dueber. Engineering Robust Control of two-component system phosphotransfer using modular scaffolds. *Proceedings of the National Academy of Sciences of the United States of America*, pages 18090–18095, October 2012.

[46] WR Whitaker. Engineering Modular Post-Translational Control Strategies in Prokaryotes. *PhD Thesis, UC Berkeley, USA*, pages 1–151, June 2012.

[47] Y Zhu and M Inouye. The role of the G2 box, a conserved motif in the histidine kinase superfamily, in modulating the function of EnvZ. *Molecular Microbiology*, 45(3):653–663, July 2002.

[48] CF Huang and JE Ferrell, Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proceedings of the National Academy of Sciences of the United States of America*, 93:10078–10083, September 1996.

[49] ES Groban, EJ Clarke, HM Salis, SM Miller, and CA Voigt. Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors. *Journal of Molecular Biology*, 390(3):380–393, July 2009.

[50] B Munsky, B Trinh, and M Khammash. Listening to the noise: random fluctuations reveal gene network parameters. *Molecular Systems Biology*, 5, October 2009.

[51] L Jiang, EA Althoff, FR Clemente, L Doyle, D Röthlisberger, A Zanghellini, JL Gallaher, JL Betker, F Tanaka, CF Barbas, D Hilvert, KN Houk, BL Stoddard, and D Baker. De novo computational design of retro-aldol enzymes. *Science*, 319(5868):1387–1391, March 2008.

[52] D Röthlisberger, O Khersonsky, AM Wollacott, L Jiang, J DeChancie, J Betker, JL Gallaher, EA Althoff, A Zanghellini, O Dym, S Albeck, KN Houk, DS Tawfik, and D Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453(7192):190–U4, 2008.

[53] JB Siegel, A Zanghellini, HM Lovick, G Kiss, AR Lambert, JL St Clair, JL Gallaher, D Hilvert, MH Gelb, BL Stoddard, KN Houk, FE Michael, and D Baker. Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*, 329(5989):309–313, July 2010.

[54] HK Privett, G Kiss, TM Lee, RA Blomberg, Rand Chica, LM Thomas, D Hilvert, KN Houk, and SL Mayo. Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America*, 109(10):3790–3795, March 2012.

[55] O Wiest and KN Houk. On the Transition-State of the Chorismate-Prephenate Rearrangement. *Journal of Organic Chemistry*, 59(25):7582–7584, 1994.

[56] O Wiest and KN Houk. Stabilization of the Transition-State of the Chorismate-Prephenate Rearrangement - an Ab-Initio Study of Enzyme and Antibody Catalysis. *Journal of the American Chemical Society*, 117(47):11628–11639, 1995.

[57] BSP Araujo, T Miller, and SL Mayo. In Silico Screening of Computational Enzyme Designs. *The Journal of Physical Chemistry B*, 21(S1):132, 2012.

[58] A Vedani and DW Huhta. A new force field for modeling metalloproteins. *Journal of the American Chemical Society*, 112(12):4759–4767, 1990.

[59] BSP Araujo. A molecular dynamics approach for prescreening enzyme designs. Unpublished Candidacy Report, 2010.

[60] JK Lassila, JR Keeffe, P Kast, and SL Mayo. Exhaustive Mutagenesis of Six Secondary Active-Site Residues in Escherichia coli Chorismate Mutase Shows the Importance of Hydrophobic Side Chains and a Helix N-Capping Position for Stability and Catalysis . *Biochemistry*, 46(23):6883–6891, June 2007.

[61] JK Lassila, JR Keeffe, P Oelschlaeger, and SL Mayo. Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity. *Protein Engineering, Design & Selection : PEDS*, 18(4):161–163, April 2005.

[62] LM Guzman, D Belin, MJ Carson, and J Beckwith. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, 177(14):4121–4130, July 1995.

[63] F Jacob and J Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, 1961.

[64] GL Hazelbauer and WC Lai. Bacterial chemoreceptors: providing enhanced features to two-component signaling. *Current Opinion In Microbiology*, 13(2):124–132, 2010.

[65] GS Baird, DA Zacharias, and RY Tsien. Circular permutation and receptor insertion within green fluorescent proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 96(20):11241–11246, 1999.

[66] G Guntas, SF Mitchell, and M Ostermeier. A molecular switch created by in vitro recombination of nonhomologous genes. *Chemistry & Biology*, 11(11):1483–1487, 2004.

[67] NA Sallee, BJ Yeh, and WA Lim. Engineering modular protein interaction switches by sequence overlap. *Journal of the American Chemical Society*, 129(15):4606–4611, April 2007.

[68] O Erster, M Eisenstein, and M Liscovitch. Ligand interaction scan: a general method for engineering ligand-sensitive protein alleles. *Nature Methods*, 4(5):393–395, May 2007.

[69] O Erster, R Seger, and M Liscovitch. Ligand interaction scan (LIScan) in the study of ERK8. *Biochemical and Biophysical Research Communications*, 399(1):37–41, August 2010.

[70] KA Reynolds, RN McLaughlin, and R Ranganathan. Hot spots for allosteric regulation on protein surfaces. *Cell*, 147(7):1564 – 1575, 2011.

[71] GS Anand, PN Goudreau, JK Lewis, and AM Stoc. Evidence for phosphorylation-dependent conformational changes in methylesterase CheB. *Protein Science*, 9(5):898–906, May 2000.

[72] AJ Riemen and ML Waters. Dueling post-translational modifications trigger folding and unfolding of a beta-hairpin peptide. *Journal of the American Chemical Society*, 132(26):9007–9013, July 2010.

[73] RS Signarvic and WF DeGrado. De novo design of a molecular switch: phosphorylation-dependent association of designed peptides. *Journal of Molecular Biology*, 334(1):1–12, 2003.

[74] KE Luker, MCP Smith, GD Luker, ST Gammon, H Piwnica-Worms, and D Piwnica-Worms. Kinetics of regulated protein-protein interactions revealed with firefly luciferase complementation imaging in cells and living animals. *Proceedings of the National Academy of Sciences of the United States of America*, 101(33):12288–12293, August 2004.

[75] JA Ubersax and JE Ferrell. Mechanisms of specificity in protein phosphorylation. *Nature Reviews. Molecular Cell Biology*, 8(7):530–541, July 2007.

[76] FW Herberg, SM Bell, and SS Taylor. Expression of the catalytic subunit of camp-dependent protein kinase in escherichia coli: multiple isozymes reflect different phosphorylation states. *Protein Engineering*, 6(7):771–777, 1993.

[77] KE Thompson, CJ Bashor, WA Lim, and AE Keating. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synthetic Biology*, 1(4):118–129, April 2012.

[78] HS Park, MJ Hohn, T Umehara, LT Guo, EM Osborne, J Benner, C. Noren, J Rinehart,

and D Söll. Expanding the genetic code of escherichia coli with phosphoserine. *Science*, 333(6046):1151–1154, 2011.

[79] IU Heinemann, AJ Rovner, HR Aerni, S Rogulina, L Cheng, W Olds, JT Fischer, D Söll, FJ Isaacs, and J Rinehart. Enhanced phosphoserine insertion during escherichia coli protein synthesis via partial {UAG} codon reassignment and release factor 1 deletion. *Federation of European Biochemical Societies Letters*, 586(20):3716 – 3722, 2012.

[80] Z Songyang, S Blechner, N Hoagland, MF Hoekstra, H Piwnica-Worms, and LC Cantley. Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Current Biology*, 4(11):973 – 982, 1994.

[81] O Dagliyan, D Shirvanyants, AV Karginov, F Ding, L Fee, SN Chandrasekaran, CM Freisinger, GA Smolen, A Huttenlocher, KM Hahn, and NV Dokholyan. Rational design of a ligand-controlled protein conformational switch. *Proceedings of the National Academy of Sciences of the United States of America*, 110(17):6800–6804, 2013.

[82] A Kumar, M Zhang, L Zhu, RP Liao, C Mutai, S Hafsat, DR. Sherman, and MW Wang. High-throughput screening and sensitized bacteria identify an *M. tuberculosis* dihydrofolate reductase inhibitor with whole cell activity. *Public Library of Science ONE*, 7(6):e39961, June 2012.

[83] DE Cameron and JJ Collins. Tunable protein degradation in bacteria. *Nature Biotechnology*, 32(12):1276–1281, December 2014.

[84] A Levskaya, OD Weiner, WA Lim, and CA Voigt. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature*, 461(7266):997–1001, October 2009.