Prototyping Diverse Synthetic Biological Circuits in a Cell-Free Transcription-Translation System

Thesis by

Shaobin Guo

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

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
(Defended May 10, 2017)

Shaobin Guo
ORCID: 0000-0001-9736-4078

ACKNOWLEDGEMENTS

It has been a great experience to study here at California Institute of Technology for the past 6 years (2011-2017). I cannot be more grateful for the opportunity to come from Xiamen University, China to the United States to pursue my doctorate degree. Not only do I expand my knowledge in biological sciences, but I also learn to live in a foreign country, speak a different language, appreciate the similarities and differences between China and US.

First, I would like to thank my PhD advisor Prof. Richard Murray for his tremendous support and mentorship for the past 5 years. It has been a great pleasure to work with him and learn from him. He left me room to work and think independently but was always there when I needed help and guidance. During some of the most difficult times here at Caltech, it was his encouragement and comfort that carried me through the challenges.

I would also like to extend my acknowledgements to my thesis committee members, Prof. Michael Elowitz, Lea Goentoro, and Paul Sternberg for their invaluable support and advice, and all the other professors, who encouraged and helped me during my PhD time here, Prof. Doug Rees, Bil Clemons, Shu-ou Shan, and Jackie Barton.

Being part of the Murray lab is a wonderful experience, because of the people here. I would like to thank my previous and current colleagues for being so helpful and understandable. It was always a mixed feeling when fellow Murray lab members graduated and left us for good, and sadly it is now my turn to say goodbye to the lab, to the lab mates, and to the friends.

Lastly, I would like to thank my family for their unconditional love. To my parents: Thank you for bringing me to this world and raising me up to be the person I am today. Thank you for the daily video chats that accompany me when I miss home. I am so fortunate to have you in my life and there is no word that can fully express my gratitude to both of you. I love you. 致我的父母亲：感谢您们带我来到这个世界并将我抚养长大。感谢您们每天跨洋视频的陪伴，抚慰我的思乡之情。能有您们在我的生命中，我感到无比幸福，千言万语尽在不言中。我爱您们。

To my wife, Cecilia: I did not believe in fate and destiny until I met you. Although we both came from Fuzhou, China, we did not know each other until we met here in Los Angeles. I am so lucky to fall in love with you and get married to you. Thank you for bringing our daughter to the world. Because of you, I have had so much more fun in life than ever before. I am looking forward to all the new adventures that we are going to experience together. Like the saying goes: life is great now, but the best is yet to come. I love you.

To my daughter, Rory: My little girl, you are the greatest gift ever in my life. It is you who reminds me that life can be full of surprise and happiness. It is always a dilemma: I cannot wait for you to grow up and be able to read this (and my research) but I am afraid that I am going to miss the little chubby baby that you are now. I love you.

ABSTRACT

Synthetic biological circuits are the foundation for the ultimate goals of controlling cells and building artificial cells from the ground up. To get closer to these goals in a more efficient way, we utilize a cell-free transcription-translation system to help perfect biological circuits for the simplicity, freedom, and convenience that the system offers. In this thesis, we demonstrate three distinct aspects of biological circuits in a cell-free transcription-translation system: circuit dynamics, phosphorylation, and membrane proteins. We start with a simple feedforward circuit, which shows dynamic responses to the input. We first prototype the feedforward circuit in the cell-free system with the aid of mathematical modeling. Then, based on the knowledge learned from prototyping, we successfully implement the circuit in cells. Not only do we show that a circuit with dynamics can be prototyped in the cell-free system, but we also test a more complicated circuit involving a phosphorylation cycle. The phosphorylation-based insulator circuit is prototyped and then a model created for the circuit is shown to be identifiable in the cell-free system. To further expand the capability of the cell-free system, we demonstrate that biologically active membrane proteins can be generated in the cell-free system with engineering, suggesting that even biological circuits requiring membrane proteins can be prototyped in the system. These results help advance our knowledge of both biological circuits and the cell-free transcription-translation system, and bring us one step closer to our ultimate goals of implementing control theory in synthetic biology.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] Guo, S. and R.M. Murray, *Prototyping And Implementation Of A Novel Feedforward Loop In A Cell-Free Transcription-Translation System And Cells.* bioRxiv, 2017. DOI: https://doi.org/10.1101/123190.

S.G. and R.M.M. conceived the idea. S.G. designed the experiments, performed the experiments and subsequent data analysis, and wrote the manuscript. R.M.M provided comments on the manuscript.

[2] Guo, S., E. Yeung, and R.M. Murray, *Implementation and System Identification of a Phosphorylation-Based Insulator in a Cell-Free Transcription-Translation System.* bioRxiv, 2017. DOI: https://doi.org/10.1101/122606. (Submitted)

S.G. and E.Y. contributed equally to this work. S.G., E.Y. and R.M.M. conceived the idea. S.G. and E.Y. designed the experiments, performed the experiments and subsequent data analysis and wrote the manuscript. R.M.M. provided comments and revised the manuscript.

[3] Guo, S., A. Vaish, Q. Chen, and R.M. Murray, *Expressing Biologically Active Membrane Proteins in a Cell-Free Transcription-Translation Platform.* bioRxiv, 2017. DOI: https://doi.org/10.1101/104455. (Undergoing internal review)

S.G., A.M., Q.C. and R.M.M. conceived the idea. S.G. designed the experiments, performed the experiments and subsequent data analysis and wrote the manuscript. A.M. participated in the design and execution experiments. A.M., Q.C. and R.M.M. provided comments and revised the manuscript.

TABLE OF CONTENTS

Acknowledgements……...iii
Abstract ……iv
Published Content and Contributions…….....v
Table of Contents…… vi
List of Illustrations and/or Tables……vii
Introduction 1
Chapter 1: Prototyping and implementation of a novel feedforward loop in a cell-free transcription-translation system and cells.....3
1.1 Introduction 4
1.2 Circuit design and simulations.....6
1.3 Linear DNAs and plasmids construction and prototyping in TX-TL 9
1.4 Implementation of the biocircuit in cells 11
1.5 Discussion 12
Materials and Methods.....15
Supplementary Materials
Implementation of the feedforward loop composed of LasR and pLas 18
Chapter 2: Implementation and system identification of a phosphorylation based insulator in a cell-free transcription-translation system 23
2.1 Introduction 25
2.2 Demonstration of Retroactivity in TX-TL.....28
2.3 Demonstration of the Insulation Capability of the PBI Circuit.....30
2.4 Estimation of Constitutively Expressed Protein Concentrations 34
2.5 Derivation of A Simplified Model for the PBI.....45
2.6 System Identification of the Simplified PBI Model 51
2.7 Simulations.....60
2.8 Conclusion.....62
Materials and Methods.....64
Supplementary Materials
Supplementary Figures 66
Brief introduction on retroactivity and how to attenuate it 69
Chapter 3: Expressing biologically active membrane proteins in a cell-free transcription-translation platform.....74
3.1 Introduction 76
3.2 Results and Discussion.....78
3.3 Conclusions 87
Materials and Methods.....89
Supplementary Materials
Supplementary Figures 92
SM1 The trials and errors of the expressions of β1AR/ β2AR proteins in TX-TL 94

SM2 Expressing a chimera potassium channel membrane protein KcsA-Kv1.3 in TX-TL 108
Appendix I: Quantification of Terminator Strengths Using Linear DNAs in TX-TL.....115
Conclusion and Future Directions.....120
Bibliography 122

LIST OF ILLUSTRATIONS AND/OR TABLES

Page

Chapter 1

Figure 1: Diagrams of feedforward loops ... 5
Figure 2: Simulations of the FFL ... 8
Figure 3: Experimental results of the FFL ... 10
Figure S1: Illustration diagram of the FFL composed of LasR ... 18
Figure S2: Time course data from the *in vivo* experiment of the LasR FFL ... 19
Figure S3: Time course of the selected pLas promoter ... 20
Figure S4: Time course data from the experiment of the optimized LasR FFL ... 21
Figure S5: Microscope movie snapshots of the optimized LasR FFL ... 22

Chapter 2

Figure 1: Demonstration of the retroactivity in the TX-TL system ... 29
Figure 2: Implementation of the PBI circuit in the TX-TL system ... 31
Figure 3: Data featuring mSpinach expression on linear DNA ... 40
Figure 4: Fitting plots of $dGFP/dt$ against pGlnA and $P\mathrm{h}/K$ ... 59
Figure 5: Simulation results of the PBI circuit based on system identification ... 61
Figure S1: Demonstration of the phosphorylation cycle in TX-TL ... 66
Figure S2: Test the temperature sensitivity of the PBI circuit ... 67
Figure S3: Linear regression to estimate the GFP production rate ... 68
Figure S4: A schematic illustrating the structure and function of the PBI ... 70

Chapter 3

Table 1: Information of GPCR constructs used in experiments ... 78
Figure 1: Illustration of the TX-TL experimental setup ... 79
Figure 2: Expression of pSG73-75 constructs in TX-TL ... 81
Figure 3: Western blots of TX-TL reactions ... 83
Figure 4: Binding assays of TX-TL made β1 and β2 adrenergic receptor ... 86
Figure S1: Bar chart showing the impact of various detergents on TX-TL ... 92
Figure S2: Bar chart of carazolol fluorescence fold change ... 93
Figure S3: Illustration diagram of linear DNAs used for expression in TX-TL ... 95
Figure S4: Bar chart of the expression from constructs of interest ... 96
Figure S5: Gel results of the first test ... 97
Figure S6: Bar chart of the expression from constructs of interest in 2nd test ... 98
Figure S7: Gel results of the 2nd test ... 100
Figure S8: Bar chart of the expression from constructs of interest in 3rd test ... 102

Figure S9: Gel results of the 3rd test 105
Figure S10: Western blot results of fusion proteins with reducing agents 107
Figure S11: Gel results from Kv1.3 expression 109
Figure S12: Western blot results of pSG66 and pSG69 expressed in TX-TL 111
Figure S13: Western blot results of refolded Kv1.3 113
Figure S14: Kv1.3 binding assay 114
Appendix I .....
Figure 1: Illustration diagrams of constructs used for terminator strength assay 116
Figure 2: The terminator strength assay data 119

INTRODUCTION

The relatively recent rise of synthetic biology mainly begins with the publications of a synthetic repressilator [4] and a genetic toggle switch [5]. As more efficient biotechnology tools are being developed [6-10], the young field has advanced significantly [11, 12]. In this thesis, we focus on one important aspect of the synthetic biology – synthetic biological circuits. Biological circuits, the basis of synthetic biology, can be used to program cells to perform distinct functions, such as biosensors [13, 14], metabolites synthesis [15-17], and so on. However, the lack of modularity and robustness of some biological circuits often limits their applications [18, 19]. In order to deepen our understanding of biological circuits in a more efficient fashion, we utilize a cell-free transcriptional-translation system (TX-TL) to prototype, characterize, and implement the biocircuits [20, 21].

In Chapter 1, we demonstrate that a simple novel synthetic biological circuit, exhibiting dynamic response, can be prototyped in the TX-TL system with the aid of mathematical modeling. We further show that the characterized feedforward loop circuit can be implemented in cells with minimum optimizations. This chapter, together with previous studies [22-24], suggests that the TX-TL system is an easy-to-use, convenient, and efficient approach to prototype biocircuits.

In Chapter 2, built on our understanding from chapter 1, we prototype and characterize a more complicated biocircuit – a phosphorylation-based insulator – in the TX-TL system. We demonstrate that non-transcriptional regulations, such as a phosphorylation cycle, can be prototyped in the TX-TL system as well. Not without

its limitations, but TX-TL, as shown in this chapter, provides a useful system for not only circuit prototyping but also system identification. With the additional degrees of freedom provided by the TX-TL system (comparing to cell-based systems), we can identify the parameters of our mathematical model for this biocircuit using actual experimental data.

Finally, in Chapter 3, we discuss the possibility of expanding the capabilities of the TX-TL system. We demonstrate that biologically active membrane proteins can be expressed in the membrane-less TX-TL system with further engineering. Membrane proteins, which consist of many important receptors and enzymes, will be valuable additions to the current biocircuit component library. The ability to prototype biocircuits involving membrane proteins in the TX-TL system will also be very beneficial to the field.

Supplementary materials included in this thesis also provide helpful insights into our progress of prototyping biocircuits in TX-TL. Specifically: we successfully prototype additional feedforward loop circuit in the TX-TL; we characterize the strengths of a group of genetic terminators on linear DNAs in the TX-TL; and we show that TX-TL made membrane proteins can be refolded into correct conformations.

Chapter 1

PROTOTYPING AND IMPLEMENTATION OF A NOVEL FEEDFORWARD LOOP IN A CELL-FREE TRANSCRIPTION-TRANSLATION SYSTEM AND CELLS

Abstract

Building novel synthetic biological devices is a time-consuming task because of lengthy cell-based testing and optimization processes. Recent progress made in the cell-free field suggests that the utilization of mathematical models and cell-free transcription-translation testing platforms to systematically design and test novel synthetic biocircuits may help streamline some of the processes. Here we present a study of building a novel functional biological network motif from scratch with the aid of the mathematical modeling and the cell-free prototyping. In this work, we demonstrate that we were able to make a 3-promoter feedforward circuit from a concept to a working biocircuit in cells within a month. We started with performing simulations with a cell-free transcription-translation simulation toolbox. After verifying the feasibility of the circuit design, we used a fast assembling method to build the constructs and used the linear DNAs directly in the cell-free system for prototyping. After additional tests and assemblies, we implemented the circuit in plasmid forms in cells and showed that the *in vivo* results were consistent with the simulations and the outcomes in the cell-free platform. This study showed the usefulness of modeling and prototyping in building synthetic biocircuits and that we can use these tools to help streamline the process of circuit optimizations in future studies.

1.1 Introduction

Traditional methods for building synthetic biological circuits are labor-intensive and time-consuming [25]. Recent research progress on a cell-free bimolecular breadboard platform provides us with a potential tool to perform fast circuit prototyping *in vitro* [22-24]. The *Escherichia coli*-based cell-free transcription-translation system (TX-TL) is a "biomolecular breadboard" that allows us to quickly design, build, test, and debug novel synthetic biocircuits *in vitro* [20, 21]. Like a wind tunnel is to airplanes or a breadboard is to electronic circuits, TX-TL allows a biological engineer to quickly test, debug, and retest their biological circuits *in vitro*, bypassing the time-consuming steps of cloning, transformation, and cell growth, which are required for *in vivo* testing.

TX-TL is a cell-free system based on S30 cell extracts. The extracts have been optimized for *in vitro* biocircuits testing, which means it mimics the *E. coli in vivo* characteristics while preserving transcription, protein production capability, and regulatory mechanisms [20, 21]. Previous work has shown that, besides plasmids, linear DNAs can also be used in TX-TL for fast circuit prototyping with the protection from the RecBCD inhibitor bacteriophage gamS protein [22, 26]. Combine this with the mathematical modeling, and it is possible to rapidly design and characterize functional synthetic biocircuit modules, such as the gene regulation network motifs, in TX-TL.

Gene regulation networks are composed of a small set of recurring interaction

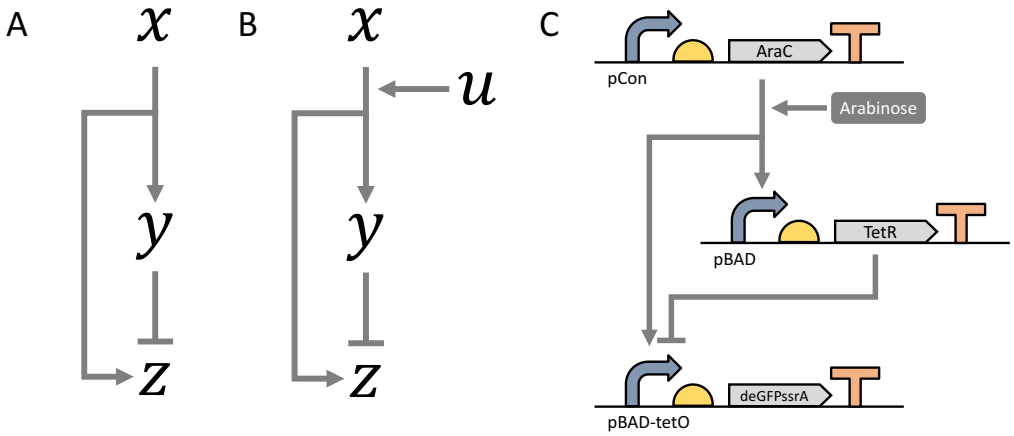


Figure 1: Diagrams of feedforward loops. **A:** The original feedforward loop illustration, composing of three components x, y and z. Arrows mean activation and bars mean repression. **B:** The adapted feedforward loop illustration, with additional input u to control the start of the activation. **C:** The actual biocircuit design for the feedforward loop. The activator AraC is at the 3' of a constitutive promoter pCon. Arabinose is the inducer input. The repressor is controlled by the promoter pBAD, which is regulated by AraC/arabinose. The reporter deGFPssrA is at the 3' of a combinatorial promoter pBAD-tetO, which can be activated by AraC/arabinose but can also be repressed by TetR.

patterns called network motifs [27, 28]. Among all the common gene regulation network motifs, the incoherent type-1 feedforward loop (FFL) is one of the widely used and interesting ones [29]. This specific FFL is composed of two input transcription factors, one (x) of which activates the other (y), and both jointly regulate (activate or repress) a target gene (z) (Figure 1A). Because there is a time delay between the activation of z by x and the repression of z by y, it has been shown that this type of FFL can generate a temporal pulse of z response.

Using this FFL as an example, here we demonstrate the process of using a mathematical model and the TX-TL prototyping platform to build novel synthetic biocircuits and characterize their behaviors. First, we verified our circuit design by

performing simulations in a cell-free transcription-translation toolbox (TX-TL toolbox) [30]. Next, we built the constructs based on the verified design using a fast assembly method [31] and then we used the linear DNAs of these constructs directly for test in the TX-TL [21, 22]. After additional tests and assemblies, we implemented the FFL circuit *in vivo* and saw consistent results as *in silico* and *in vitro*. This study brought attention to utilizing mathematical models and the TX-TL prototyping platform when designing novel synthetic biocircuits. Instead of building and testing circuits directly in cells, we can save significant amounts of man-hours and streamline some of the prototyping steps involved in building new network motifs by properly employing this TX-TL system.

1.2 Circuit design and simulations

To make things simpler in both TX-TL and subsequent cell-based tests, we first optimized the circuit design. Instead of having x directly turning on y and z, we added another component (u) to the circuit for better control (Figure 1B). For TX-TL and especially cell-based tests, we will put all three components of the circuit (x, y, and z) into the testing platforms in the beginning. In order to control when the circuit should start the dynamics, an extra input is required. Without the presence of the input u, x cannot activate any of the downstream components. But as soon as u is added, the complex $[x{:}\,u]$ will activate both y and z and initiate the dynamics. In this case, the activation of y and z is positively correlated with u, and u can be seen as an inducer that is required for x to activate the downstream parts.

Having the basic design in mind, we looked for biological parts that would fit the requirements stated above. We decided to use the AraC-arabinose activation system as $[x:u]$ [32], a transcription factor TetR as the repressor y [33] and deGFP fluorescent protein with the corresponding combinatorial promoter as output z [34] (Figure 1C). The transcription factor AraC binds to the promoter pBAD and activates the transcription of downstream genes (TetR and deGFP) only in the presence of the inducer arabinose. On the other hand, the transcription factor TetR binds to operator site tet and represses the transcription of deGFP. A small molecule anhydrotetracycline (aTc) can be used to sequester TetR proteins away from the operator and resume the transcription. Because of the time delay between the activation by AraC-arabinose and the repression by TetR, the deGFP gene will first be transcribed, and only after TetR proteins accumulate to certain threshold level (aTc can be used to extend the delay), the transcription of deGFP ceases. At the same time, the degradation ssrA tag on the deGFP protein leads to degradation of the protein [35]. As a result, the green fluorescence signal, which can be measured, will first increase and then decrease, generating a pulse-like behavior (Figure 2A).

Simulations were performed using the TX-TL toolbox [30]. By tuning the parameters in the toolbox model, we verified the feasibility of our design through simulations. In Figure 2A, we can see that the model captures the pulse-like behavior of deGFP protein concentration, along with the increasing activator AraC

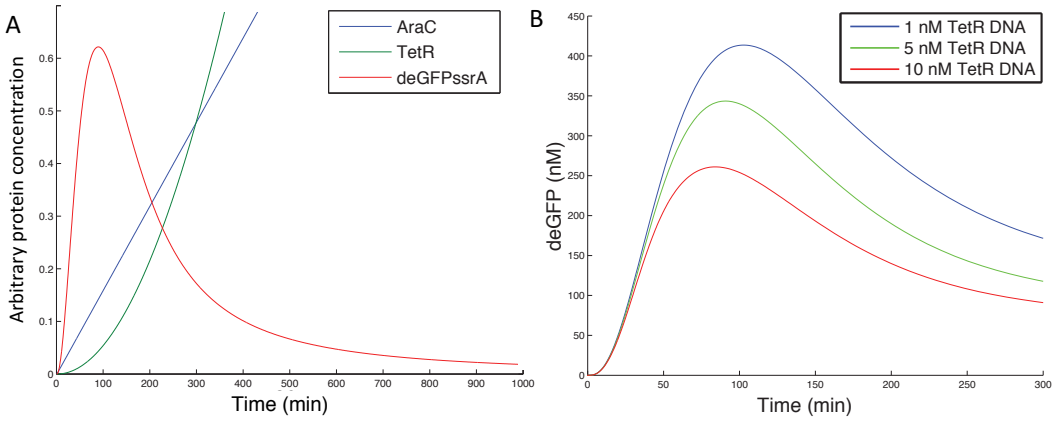


Figure 2: Simulation results of the FFL using the TX-TL simulation toolbox. **A:** Simulation of the time course of the arbitrary protein concentrations for all three circuit components. Initial AraC, TetR and deGFP DNA concentrations are set to be the same at 1. Arabinose concentration is also set to 1. AraC, which is controlled by a constitutive promoter, is assumed to be produced in a linear fashion. The pBAD promoter controlling deGFP is set to be stronger than the pBAD promoter controlling TetR, in order to extend the delay and produce significant amount of deGFP proteins. **B:** Simulation of the time course of the deGFP protein concentrations with varied initial TetR DNA concentrations from 1 nM to 10 nM. Initial conditions for AraC and deGFP were 10 nM and the arabinose concentration was set to be 0.2%.

concentration and repressor TetR concentration. The delay in the increase of TetR concentration and the degradation of deGFP proteins result in the pulse-like behavior in deGFP concentration.

In both the mathematical model and the TX-TL platform, we have fully control over the initial DNA concentrations for all the components. That gives us the freedom to change the circuit dynamics by simply changing the inputs – DNA concentrations. In Figure 2B, we tested how the varied concentrations of the repressor TetR DNA can affect the deGFP dynamics with all the other components as constants. As the simulation result shows here, increasing TetR DNA concentrations not only brings down the peak deGFP concentrations, but also shifts the peak to the left, suggesting

that it takes less time for TetR protein to accumulate to the threshold level when there is more initial TetR DNA.

1.3 Linear DNAs and plasmids construction and prototyping in TX-TL

After verifying the circuit design using simulations, we decided to build the incoherent type-1 FFL shown in Figure 1C. We first designed and ordered primers (a day before) to amplify the coding sequences for AraC, TetR, and deGFPssrA. Then we used the GoldenBraid assembly method to stitch specific promoters, ribosome binding sites (RBSs), coding sequences (CDSs) and terminators together with plasmid vectors [31]. After 1 hour incubation, we amplified the linear DNAs containing Promoter-RBS-CDS-Terminator-Vector via PCR reactions. Then we used these linear DNAs to run experiments in TX-TL directly with the presence of gamS. From start to finish, one experiment cycle can be less than a day.

TX-TL experiments were run by simply mixing the extract and buffer with the inputs – DNAs (AraC, TetR and deGFP) and the inducer arabinose [21]. After mixing all together, GFP fluorescence was measured using a plate reader. Figure 3A showed the experimental results from TX-TL experiments. All the curves were consistent with the simulation: GFP signal first increased as a result of AraC-arabinose activation; then after TetR proteins accumulated to the threshold amount, they repressed the transcription of deGFPssrA and at the same time, ClpX protein, which is an ATPase, unfolded the tagged deGFP proteins and caused the reduction of GFP signal [36]. As we can also see in the figure, the more the TetR DNAs were added,

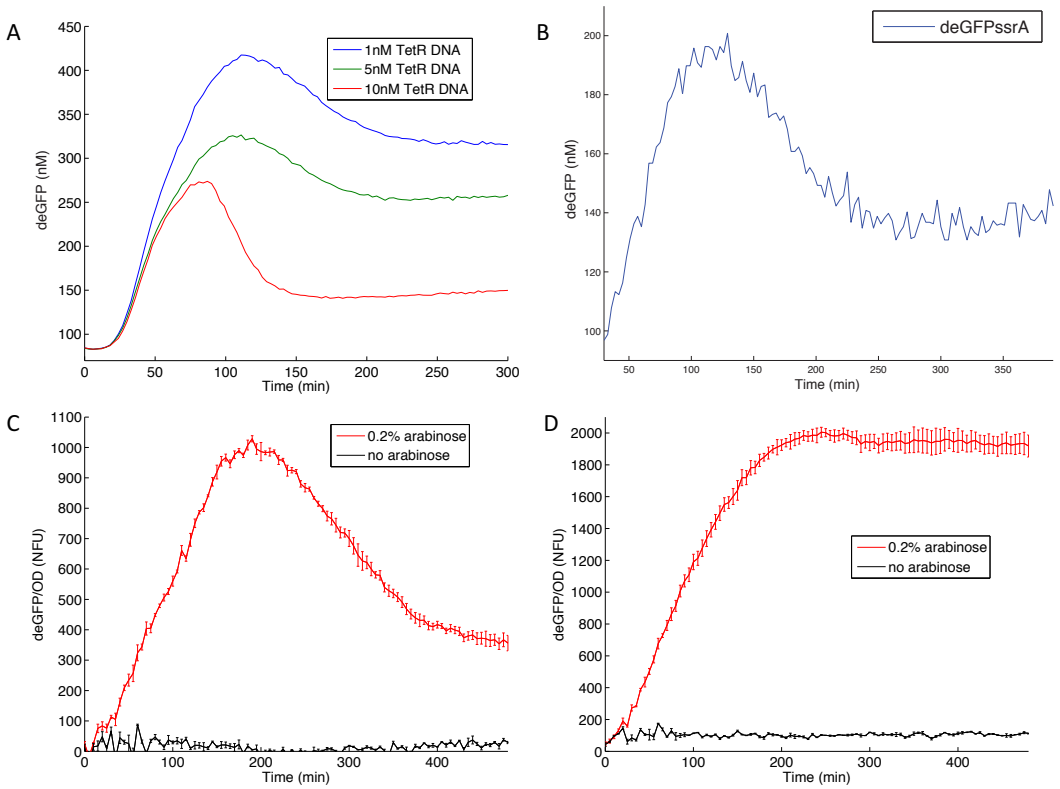


Figure 3: Experimental results of the FFL in TX-TL and cells. **A:** The time course result of the FFL experiment in TX-TL with linear DNA: 10 nM AraC linear DNA, 10 nM deGFP linear DNA, 10 nM ClpX linear DNA, 0.2% arabinose, 0.1 µg/mL aTc and varied TetR linear DNA concentrations. Experiments were run at 29°C. **B:** The time course result of the FFL experiment in TX-TL with plasmid DNA: 2 nM AraC-TetR-deGFPssrA plasmid, 1 unit of purified ClpX protein, 0.2% arabinose and 0.01 µg/mL aTc. Experiments were run at 29°C. **C, D:** The time course results of the FFL experiments in cells. **C:** deGFP protein is tagged with ssrA degradation tag. **D:** deGFP does not have a degradation tag. GFP signals, which were average from three independent repeated wells, were measured using a plate reader and then data were normalized using OD600 readings to get the fluorescence reading for each cell. The concentration of the inducer arabinose is 0 or 0.2%. Experiments were run at 37°C.

the lower the GFP signal was and the faster the signal reached peak; this is also consistent with our simulations. However, it is easy to notice that the steady state GFP protein concentrations for different initial TetR DNA concentrations were different. This is because TX-TL reactions have limited resources, including RNA polymerases, NTPs, ribosomes, and amino acids [37]. The energy required for ClpX

to unfold tagged deGFP proteins will run out gradually, and the resources will get used up faster when more DNAs are in the reaction. As a consequence, the GFP concentrations showed different steady state levels.

After we tested the linear DNA version of the FFL in TX-TL and found a working design, we assembled these linear DNAs into one plasmid using GoldenBraid assembly method in order to implement the circuit in cells. Before testing the plasmid version of FFL *in vivo*, we first evaluated the construct in TX-TL, as it is fast and convenient to set up TX-TL experiments and there is no need for the time-consuming step of growing cells. Figure 3B showed the results of the plasmid version of the FFL in TX-TL. The dynamics of the circuit were again consistent with those from both simulations and linear DNA circuit, suggesting that this specific circuit design had a good chance to work in cells.

1.4 Implementation of the biocircuit in cells

Following the test of expressing plasmid version of the FFL in TX-TL, we transformed that plasmid into *E. coli* cells. In addition, to make sure the decrease of GFP signal *in vivo* is specific to the GFP degradation by ClpX instead of the dilution introduced by cell division, we had a control circuit, in which the deGFP was not tagged with ssrA degradation tag. Figure 3 showed the experimental results from *in vivo* experiments of the FFL circuit (Figure 3C) and the control circuit (Figure 3D). The dynamics shown in Figure 3C were clearly consistent with those shown in TX-TL experiments, meaning the FFL prototyped in TX-TL indeed showed the same

behavior *in vivo*. In contrast, the control circuit in Figure 3D did not exhibit a pulse-like behavior, suggesting that the pulse we saw in Figure 3C was not a result of dilution but was caused by the degradation of the GFP proteins.

1.5 Discussion

The idea of the cell-free TX-TL platform is not new, as there have been many studies on the original S30 cell extract since it first came out in the 1960s [38, 39]. However, most of the cell-free extracts were focusing on protein expression as an alternative option to making proteins in cells [40]. This specific TX-TL platform was developed with the goal of prototyping synthetic biocircuits in mind. TX-TL, along with the mathematical toolbox developed for it, can be the prototyping breadboard for synthetic biology. This work, together with other publications [20, 22-24, 37], serves as a testimony for that ambition. Expressing transcription factors and have them turning certain components on or off has been demonstrated before; but prototyping a synthetic biocircuit with spontaneous dynamics built in, such as an incoherent type-1 feedforward loop, is challenging. By tuning the parameters used by the mathematical model, we could get a sense of what strength of promoters and ribosome binding sites should be used in the actual biocircuit. Then we had to spend some amount of time characterizing specific parts, such as promoters, ribosome binding sites and coding sequences individually in TX-TL. From this preliminary parts library, then we could assemble our constructs and test the actual components in TX-TL. Because of the fast iteration time of this platform, we were able to finish all the above and the final implementation in cells within a month. Compared to the

first generation of biocircuits, we have shortened the development time significantly [4, 5]. We have also optimized a variant FFL circuit composed of a different activator in TX-TL and in cells (see "*Implementation of the feedforward loop composed of LasR and pLas*" in Supplementary Materials for details).

There are limitations to this protocol. First, there are limited resources in the TX-TL reactions and on top of that, there is competition of different components for the same transcription and translation machineries. One way to avoid this limitation is to set up reactions in compartments that only allow small molecules exchanges so that some resources, such as amino acids and NTPs, can be replenished. Second, it is challenging to use TX-TL data quantitatively to deduce the results in cells. Cell-free systems, no matter how we justify it, are different from cells. When TX-TL is used to help with prototyping novel synthetic biocircuits, it is recommended that the results are examined qualitatively instead of quantitatively. Although the absolute strengths of certain promoters and RBSs are different between TX-TL and cells, their relative strengths are comparable between the two systems. Third, the TX-TL simulation toolbox has its uncertainty and arbitrariness. The parameters used in the toolbox might not be physiologically reasonable despite the fact that we tried to refer to as many literature available parameters as possible. However, qualitatively we can use the simulation results as a reference; for example, we need one promoter to be stronger than the other promoter in order to achieve the desired dynamics and then we can have this information in mind when we design the actual biocircuits. In summary, though it is not a platform without its limitations, TX-TL could certainly

be used for rapid preliminary characterization and prototyping of synthetic biocircuits.

Materials and Methods

Cell-free experiment preparation and execution

Preparation of the cell-free TX-TL expression system was done according to previously described protocols [21], resulting in extract with conditions: 8.9 – 9.9 mg/mL protein, 5 mM Mg-glutamate, 40 mM K-glutamate, 1.5 mM each amino acid except leucine, 1.25 mM leucine, 50 mM HEPES, 1.5 mM ATP and GTP, 0.9 mM CTP and UTP, 0.2 mg/mL tRNA, 0.26 mM CoA, 0.33 mM NAD, 0.75 mM cAMP, 0.068 mM folinic acid, 1 mM spermidine, 30 mM 3-PGA, 2% PEG-8000.

TX-TL reactions were conducted in a volume of 10 μL in a 384-well plate (Nunc MicroWell 384-well optical bottom plates) at 29°C, using a three-tube system: extract, buffer, and DNA. When possible, inducers such as arabinose or purified proteins such as gamS [26] were added to a mix of extract and buffer to ensure uniform distribution. For deGFP, samples were read in a Synergy H1 plate reader (Biotek) using settings for excitation/emission: 485 nm/525 nm, gain 61 or 100. All samples were read in the same plate reader, and for deGFP relative fluorescent units were converted to either nM (for TX-TL) or Normalized Fluorescent Unit (NFU for *in vivo*) using a purified deGFP-His6 standard to eliminate machine to machine variation (different Bioteks).

PCR product preparation and plasmid DNA assembly

Linear DNA fragments were amplified using Pfu Phusion Polymerase (New England Biolabs), DpnI digested for 5 min at 37°C (New England Biolabs) while verified with agarose gel electrophoresis, and PCR purified using previously described procedures. Fragments were then assembled *in vitro* using Golden Gate assembly. For Golden Gate assembly, a 15 μL reaction was set up consisting of equimolar amounts of vector and insert, 1.5 μL 10X NEB T4 Buffer (New England Biolabs), 1.5 μL 10X BSA (New England Biolabs), 1 μL BsaI (New England Biolabs), and 1 μL T4 Ligase at 2 million units/mL (New England Biolabs). Reactions were run in a thermocycler at 10 cycles of 2 min/37°C, 3 min/20°C, 1 cycle 5 min/50°C, 5 min/80°C. For Golden Gate assembly, constructs with internal BsaI cut sites were silently mutated beforehand using a QuikChange Lightning Multi Site-Directed Mutagenesis kit (Agilent). For both methods, assembled circular DNAs were transformed into electrocompetent or chemically competent cells: a KL740 strain (lab made competent strain) if using an OR2-OR1 promoter (29°C), a MG1655 strain (lab made competent cells) for circuit testing, and a JM109 strain (Zymo Research) for all other constructs. KL740 upregulates a temperature sensitive lambda cI repressor. PCR products were amplified using Pfu Phusion Polymerase (New England Biolabs) for all constructs, and were DpnI digested. Plasmids were miniprepped using a Qiagen mini prep kit. All plasmids were processed at stationery phase. Before use in the cell-free reaction, both plasmids and PCR products underwent an additional PCR purification step using a QiaQuick column (Qiagen), which removed excess salt detrimental to TX-

TL, and were eluted and stored in 10 mM Tris-Cl solution, pH 8.5 at 4°C for short-term storage and −20°C for long-term storage. All the plasmids used in the work can be found on https://www.addgene.org/.

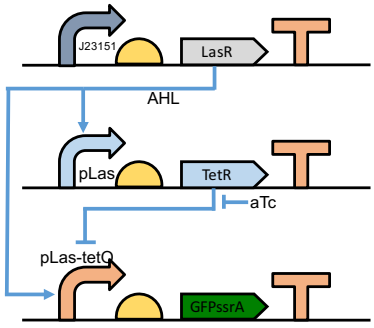
In vivo **experiment**

All *in vivo* experiments were performed in *E. coli* strain MG1655. Plasmid combinations were transformed into chemically competent *E. coli* MG1655 cells, plated on Difco LB+Agar plates containing 100 μg/mL carbenicillin and incubated overnight at 37°C. Plates were taken out of the incubator and three colonies were picked and separately inoculated into 5 mL of LB containing carbenicillin and/or chloramphenicol, and/or kanamycin at the concentrations above in a 14 mL Falcon Round-Bottom Polypropylene Tubes (Fisher Scientific), and grown approximately 17 h overnight at 37°C at 200 rpm in a benchtop shaker. This overnight culture (100 μL) was then added to a new 14mL tube containing 5 mL (1:50 dilution) of Minimal M9 casamino acid (M9CA) media [1X M9 salts (42 mM Na_2HPO_4, 24 mM KH_2PO_4, 9 mM NaCl, 19 mM NH_4Cl, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.5 μg/ml thiamine, 0.1% casamino acids, 0.4% glycerol) containing the selective antibiotics and grown for 4 h at the same conditions as the overnight culture. Then 10 μL cultures were transferred to 96-well glass bottom plate with 290 μL M9 with corresponding experimental conditions. Plates were shaken and GFP fluorescence (485 nm excitation, 525 nm emission), and optical density (OD, 600 nm) were measured using a Biotek Synergy H1m plate reader at 37°C at the highest speed for 12 hours.

Supplementary Materials

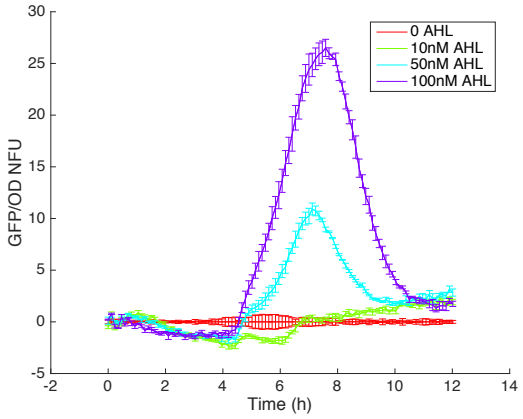
Implementation of the feedforward loop composed of LasR and pLas

Besides the feedforward loop (FFL) circuit we designed, prototyped, and implemented in the main article, we also tested a FFL designed by Zachary Sun, which had AraC, arabinose, pBAD, and pBAD-tetO replaced by LasR, N-(3-Oxododecanoyl)-L-homoserine lactone (AHL), pLas, and pLas-tetO, respectively. The circuit design is shown in Supplementary Figure S1.



Supplementary Figure S1 Illustration diagram of the FFL composed of LasR. The arrows mean activation and the bars mean repression.

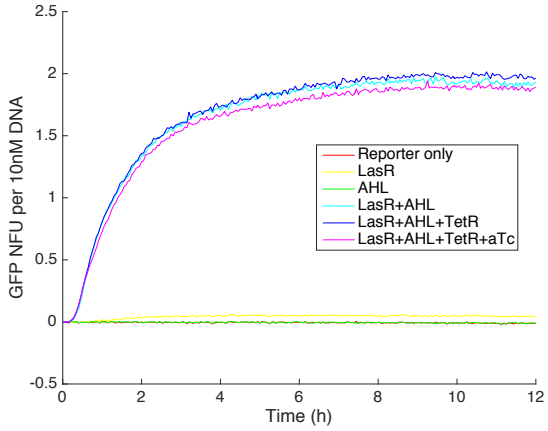
The promoter J23151 is a constitutive promoter [41]. LasR protein, in the presence of the inducer AHL, becomes an activator that can turn on both pLas and pLas-tetO promoters. All the other components work like the AraC FFL, and the time course data of the LasR circuit from *in vivo* experiments are shown in Supplementary Figure S2.



Supplementary Figure S2 Time course data from the *in vivo* experiment of the LasR FFL circuit in MG1655 *E. coli* cells. Detailed experiment setup was described in the Materials and Methods. Briefly, 4 AHL concentrations were used in the experiment, from 0 to 100 nM. 20 ng/mL of aTc was added to all of them. GFP fluorescence and OD600 measurements were done using a Biotek plate reader. GFP data was subtracted by the background and then normalized with OD data to get the normalized fluorescence unit (NFU).

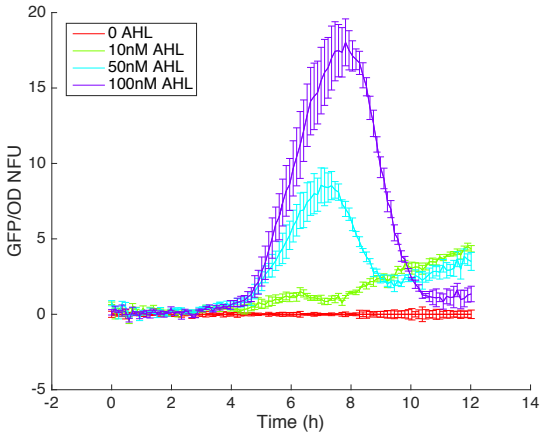
Though the LasR FFL works in a similar way as the AraC FFL, one problem with the LasR FFL circuit was that the aTc, which would bind to TetR proteins and sequester them away from tetO, had to be added to extend the delay to create significant pulses. This was due to the leaky expression of the pLas promoter. To make the circuit more robust, we engineered the pLas promoter to be more tightly controlled by LasR-AHL via prototyping different variants of pLas promoters in TX-TL. The one we found working very robustly is shown in Supplementary Figure S3 and the sequence of that pLas variant can be found on https://www.addgene.org/. As we can see in the figure, only when both LasR and AHL were added, we could see the activation of the pLas-GFP. Neither LasR nor AHL alone could activate the promoter and there was little to no leaky expression from the promoter itself. TetR

and aTc were tested to make sure that the pLas promoter, without the tetO part, could not be affected by them.



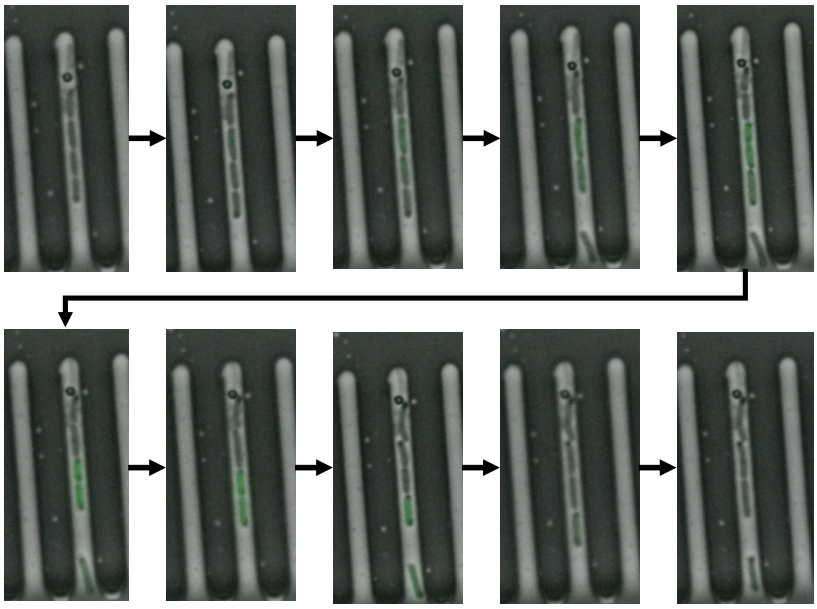
Supplementary Figure S3 Time course of the selected pLas promoter with GFP at its 3'. Experiments were run with 10 nM pLas-GFP linear DNA with or without the additional components/inducers listed in the legend. If used, the pCon-LasR linear DNA was 10 nM, pCon-TetR linear DNA was 10 nM, AHL was 50 nM and aTc was 20 ng/mL. The GFP fluorescence data was subtracted with background and then normalized using GFP protein calibration data on a Biotek plate reader.

We then modified the LasR FFL circuit with this exact promoter and we were able to generate pulse-like behavior in cells without adding any aTc. The *in vivo* data is shown in Supplementary Figure S4. As we can see, when there was no AHL added, we got no response from the circuit. Only when we had significant activation caused by AHL (more than 50 nM in this case) would we see pulse-like behavior from the circuit in cells, and the highest peaks of the pulses were positively correlated with the inducer concentrations.



Supplementary Figure S4 Time course data from the *in vivo* experiment of the optimized LasR FFL circuit in MG1655 *E. coli* cells. Detailed experiment setup was described in the Materials and Methods. Briefly, 4 AHL concentrations were used in the experiment, from 0 to 100 nM. 20 ng/mL of aTc was added to all of them. GFP fluorescence and OD600 measurements were done using a Biotek plate reader. GFP data was subtracted by the background and then normalized with OD data to get the normalized fluorescence unit (NFU).

Not only did we test the circuit quantitatively in bulks, but we also examined the circuit behavior qualitatively in a microfluidic device, also known as the mother machine. The mother machine consists of a series of growth channels that can trap single bacterial cells inside, and is designed to allow growth medium to pass through at a constant rate, which results in diffusion of fresh medium into the growth channels as well as removal of cells as they emerge from the channels into the main trench [42]. As we can see in Supplementary Figure S5, while cells were growing and dividing in a narrow comb-like space, they showed green fluorescence intensity starting from weak to strong and then back to weak with the same media keeping the inducer AHL concentration constant at 50 nM.



Supplementary Figure S5 Microscope movie snapshots of the optimized LasR FFL. Each frame was taken 10 minutes apart. Cells were grown using the same *in vivo* method described in the Materials and Methods. Fluorescence microscopy imaging was performed on an Olympus IX81 inverted fluorescence microscope using a Chroma wtGFP filter cube (450/50 BP excitation filter, 480 LP dichroic beamsplitter, and 510/50 BP emission filter), with an XFO-citep 120 PC light source at 100 % intensity and a Hamamatsu ORCA-03G camera. Cells were imaged using a 100x phase objective with oil.

Chapter 2

IMPLEMENTATION AND SYSTEM IDENTIFICATION OF A PHOSPHORYLATION-BASED INSULATOR IN A CELL-FREE TRANSCRIPTION-TRANSLATION SYSTEM

Abstract

An outstanding challenge in the design of synthetic biocircuits is the development of a robust and efficient strategy for interconnecting functional modules. Recent work demonstrated that a phosphorylation-based insulator (PBI) implementing a dual strategy of high gain and strong negative feedback can be used as a device to attenuate retroactivity. This paper describes the implementation of such a biological circuit in a cell-free transcription-translation system and the structural identifiability of the PBI in the system. We first show that the retroactivity exists in the cell-free system by testing a simple negative regulation circuit. Then we demonstrate that the PBI circuit helps attenuate the retroactivity significantly compared to the control. We consider a complex model that provides an intricate description of all chemical reactions and leveraging specific physiologically plausible assumptions. We derive a rigorous simplified model that captures the output dynamics of the PBI. We performed standard system identification analysis and determined that the model is globally identifiable with respect to three critical parameters. These three parameters are identifiable under specific experimental conditions and we performed these experiments to estimate the parameters. Our experimental results suggest that the

functional form of our simplified model is sufficient to describe the reporter dynamics and enable parameter estimation. In general, this research illustrates the utility of the cell-free expression system as an alternate platform for biocircuit implementation and system identification and it can provide interesting insights into future biological circuit designs.

2.1 Introduction

The successful design and implementation of the inaugural biocircuits, such as the genetic toggle switch and the repressilator, have demonstrated the possibility of modularity in synthetic biological circuits [4, 5]. The recognition of functional modules makes building large and complicated synthetic biological circuits possible. Basic modules can be studied and tested in isolation and then can be connected with other modules to perform certain functions. However, the modularity of biological circuits can change when interconnections are made. This effect is called retroactivity and is a fundamental issue in systems engineering [43, 44]. This means that when a downstream system is connected to another system, the downstream system will affect the behavior of upstream component. As a result, the signal generated by the upstream component may not be effectively transferred to other components.

Retroactivity can be divided into two types based on which signal it affects: the retroactivity to the input and the retroactivity to the output. Based on previous theoretical studies, an operational-amplifier-like orthogonal biomolecular device could help attenuate retroactivity [43]. An electronic operational amplifier absorbs little current from upstream; as a result, there is almost no voltage drop to upstream output. At the same time, the retroactivity to the output is attenuated because of a large amplification gain and an equally large negative feedback loop (Details are described in "*Brief introduction on retroactivity and how to attenuate it*" in Supplementary Materials). Based on these ideas and previous work [45], we tested

an insulator design using nitrogen regulation proteins [46] in a cell-free transcription-translation (TX-TL) system.

The TX-TL system developed in [20, 21] is an attractive candidate platform for such rapid prototyping. The system facilitates DNA-based expression on plasmids and linear DNA, and since linear and plasmid DNA can be prepared and expressed in the TX-TL system in a single day's time [22], the time required to iterate over designs is considerably reduced.

Another powerful aspect of the TX-TL system is the ability to directly modulate the concentration of different pieces of DNA encoding different biocircuit components. The ability to rapidly synthesize and test the effect of different promoter sites, ribosome binding sites, and other components, and simultaneously vary the DNA encoding these parts, permits a degree of freedom typically absent in cell-based assays. In this setting, iterating of prototypes can be assisted by predictive modeling of biocircuit dynamics. It is the ability to control DNA concentrations and rapidly vary structural properties of the biocircuit that allow us to address the problem of parameterizing a predictive model.

Cell-free systems have long been used to characterize fundamental parameters in biological systems [47]. In a synthetic biology context, especially for the phosphorylation-based insulator circuit, it is unclear what parametric information can be extracted from a series of systematic tests in an *in vitro* system, specifically the TX-TL system. With additional degrees of freedom in the experimental conditions,

the TX-TL system may be able to provide insight into model parameters that *in vivo* studies could not. Moreover, it is unclear what systematic tests should be carried out in order to retrieve this information. This paper investigates these issues using the phosphorylation-based insulator as a case study.

In general, a parametric model is globally structurally identifiable only under certain mathematical conditions [48]. These conditions are valid as long as the control variables enter the dynamical system as a multiplicative perturbation. However, as we will see with the phosphorylation-based insulator, even if the model retains this structure the model may not be globally identifiable because of the large number of parameters it contains, despite having only a couple output variables. As is often the case, a first principles model may be physically representative of the intricate reactions happening in the system, but carry a complexity that far exceeds the information present in the data. Thus, simplified models that are reflective of the low-dimensional output data, while also retaining the (controllable) experimental variables in the TX-TL system, are desirable.

In this work, we successfully implement the PBI circuit in TX-TL and further propose a complex model based on the fundamental processes of transcription, translation, and phosphorylation. The model is unwieldy to analyze so we rigorously derive a simplified model based on a series of physically realistic assumptions, show that it is globally identifiable with respect to the data, and perform a series of experimental perturbation tests to back out the simplified model parameters.

The main contributions of this work are the following: 1) we demonstrate that the TX-TL system can be used to prototype relatively complicated synthetic biocircuits, such as the PBI circuit that involves not only transcriptions, translations, and protein-DNA interactions but also post-translated interactions like phosphorylation and dephosphorylation; 2) we show that by utilizing the TX-TL system that has extra degrees of freedom compared to cell-based systems, we can systematically identify the parameters of our mathematical models using actual experimental data, which subsequently guide us to achieve more efficient circuit prototyping and better future circuit designs [49, 50].

2.2 Demonstration of Retroactivity in TX-TL

Firstly, we wanted to demonstrate retroactivity in the TX-TL system. The example we used is a simple negative regulation circuit, in which constitutively expressed TetR proteins repress the transcription of downstream components pTet-GFP and pTet-RFP DNA unless an inducer aTc is added (Figure 1A). Here, we considered pTet-GFP as the reporter and pTet-RFP as the load. When there is no inducer present, the reporter will remain off because of the repression by TetR. However, if we added a significant amount of load into the system, the load sequesters the TetR proteins from pTet-GFP, resulting in the activation of GFP transcription (Figure 1B). This is a result of retroactivity, in which downstream components affect the behavior of the upstream system output. We next tested this effect in the presence of different inducer concentrations (Figure 1C). At low aTc concentrations (less than 0.5 μg/mL), as load concentration increased, GFP expression increased because of retroactivity.

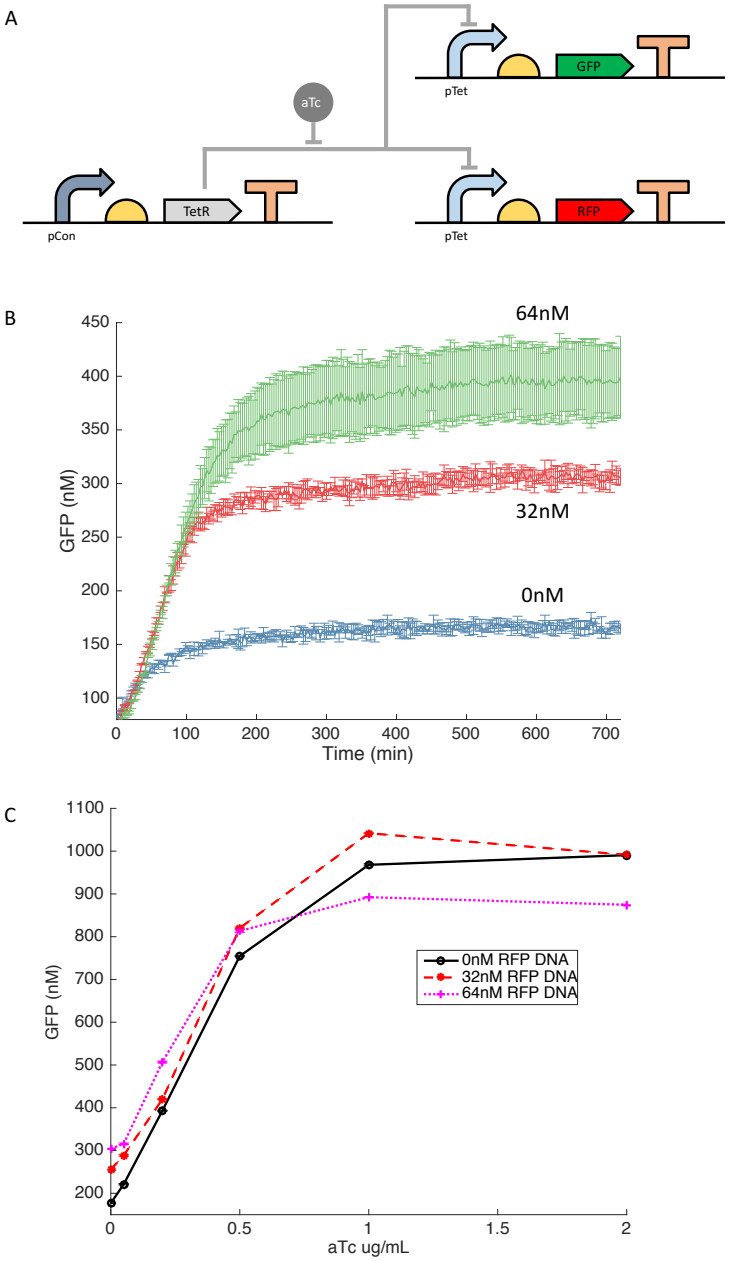


Figure 1 Demonstration of the retroactivity in the TX-TL system. **A:** Circuit diagram of a negative regulation circuit. **B:** Time traces of the GFP fluorescence in presence of different concentrations of RFP DNA. As RFP DNA concentrations increase, more GFP fluorescence can be detected as a result of the retroactivity. Error bars are standard deviations from 3 repeats. **C:** Titration of TetR repressor aTc in presence of different concentrations of RFP DNA. X axis is the final concentrations of aTc in each sample and Y axis is the end point GFP fluorescence of the corresponding samples. Data were collected using a plate reader with settings for excitation/emission: 485 nm/525 nm.

However, if too much aTc was added, GFP expression actually decreased as load increased. This is because resources in TX-TL, such as ribosomes and RNA polymerase, are limited.

This simple circuit demonstrates that there is retroactivity in biological circuits in the TX-TL system. To address this problem, we implement an insulator component to compensate for the retroactivity.

2.3 Demonstration of the Insulation Capability of the PBI Circuit

Based on the insulator design in [45], we adapted a simpler form to implement an insulator in the TX-TL system (Figure 2A). The insulator design is based on a well-known two-component signal transduction system regulating the transcription of genes encoding metabolic enzymes and permeases in response to carbon and nitrogen status in *E. coli* and related bacteria [51]. There are two essential proteins in the system: NRII and NRI (NtrB-NtrC). NRI can be phosphorylated into NRI^P by NRII (kinase form). Only NRI^P is able to activate the σ54-dependent promoter glnA and trigger the transcription of downstream genes [52]. NRII is both a kinase and phosphatase, regulated by the PII signal transduction protein, which, on binding to NRII, inhibits the kinase activity of NRII and activates the NRII phosphatase activity [53]. NRII is known to form dimers and will autophosphorylate itself to become a kinase. Previous studies suggested that when NRII has a mutation of leucine to arginine at residue 16, it loses its phosphatase activity but shows normal autophosphorylation. In contrast, NRII with a H139N mutation is not able to transfer

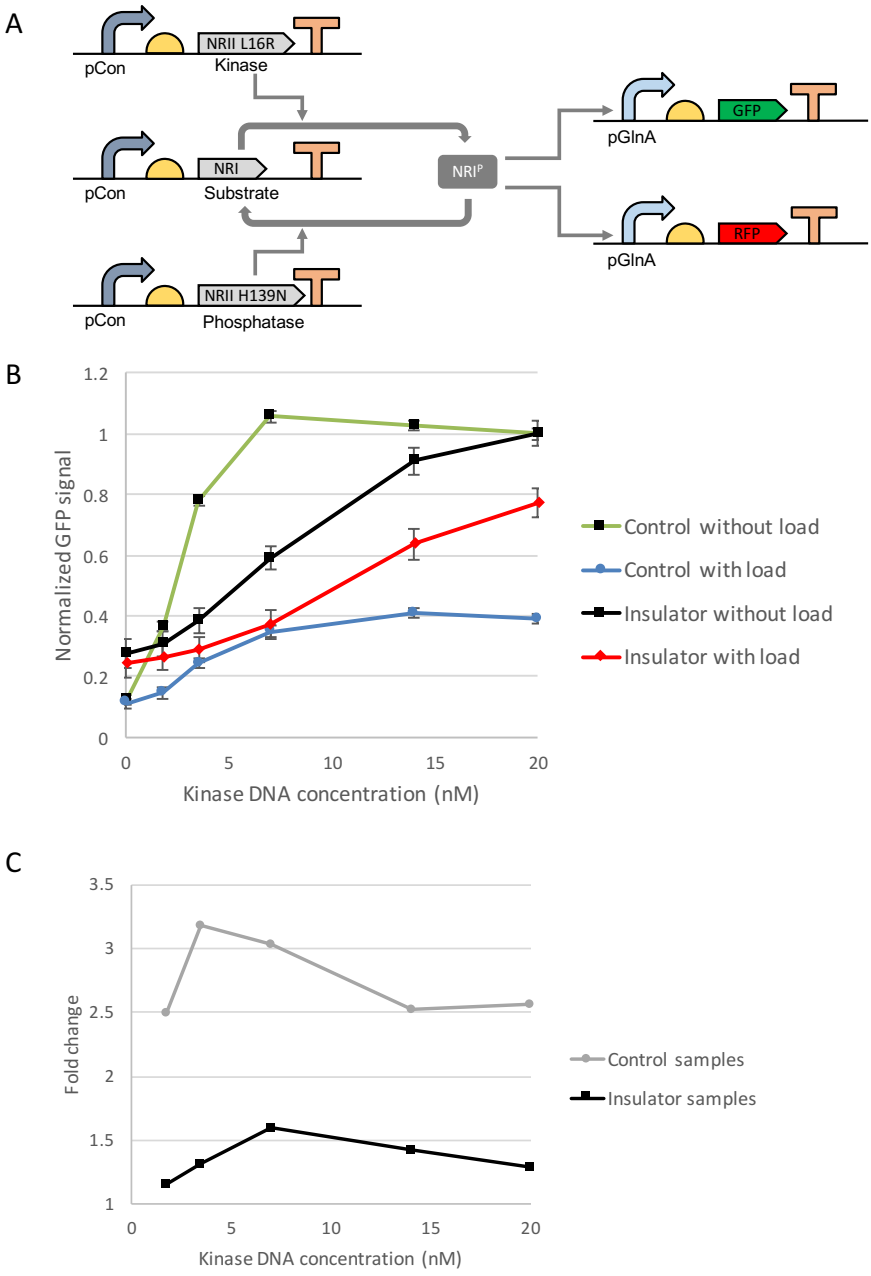


Figure 2 Implementation of the PBI circuit in the TX-TL system. **A**: Circuit diagram of the PBI circuit. pCon is a constitutive promoter. **B**: Transfer function curves for controls and insulators with or without load DNA. Raw GFP fluorescences were normalized using the highest GFP fluorescences from controls and insulators, respectively (highest GFP = 1). Compared to the control with load, which only had 40% signal left, the insulator with load was able to preserve 80% of the signal, significantly attenuating the retroactivity. **C**: Fold changes of the samples without load over the ones with load. The insulator samples have significantly smaller fold changes compared to those of the control samples.

the phosphoryl group from its active site histidine to an aspartate side chain of NRI [51]. As a result, NRIIL16R only acts as a kinase and NRIIH139N only functions as a phosphatase.

In our circuit design, proteins NRI, NRIIL16R, and NRIIH139N are all constitutively expressed. Reporter GFP is controlled by the σ54-dependent promoter glnA, which is activated by phosphorylated NRI. We also confirmed that the phosphorylation-dephosphorylation loop worked in TX-TL (Supplementary Figure S1). By virtue of the fast timescale of phosphorylation and dephosphorylation loop, this circuit enjoys a large amplification gain and an equally large negative feedback mechanism as mentioned in the introduction, which makes it a promising insulation device.

To test the insulation capability of our insulator, we compared the behaviors of the insulator circuit with a control circuit that does not have a large amplification gain, nor a negative feedback loop. As mentioned in previous theoretical studies [43], the insulator circuit requires abundant substrate NRI to achieve high gain. So we added 47.5 nM NRI linear DNA in the insulator circuit and only 5 nM in the control circuit. But to take resource limitations into account, we balanced the insulator and control by adding 42.5 nM of extra DNA (pTet-RFP) in the control circuit. Then we varied the amount of downstream glnA promoters by adding 0 nM (without-load) or 20 nM (with-load) pGlnA-RFP linear DNA, which would introduce retroactivity. We then titrated with different concentrations of NRIIL16R (kinase) linear DNA. After data was collected using a plate reader, the GFP relative

fluorescence unit of the control and insulator circuits were normalized using their highest without-load fluorescence readings, respectively. As shown in Figure 2B, when the control circuit was added with load DNA, the GFP expression dropped significantly (by about 60% at the highest kinase DNA concentration). In contrast, the insulator circuit only showed about a 20% decrease in the GFP expression when the load was added at the highest kinase DNA concentration. Figure 2C simplifies the four curves into two curves by looking at the fold changes of without-load samples over with-load samples. As we can see, the insulator samples have much smaller fold change between without-load and with-load samples compared to the control samples, indicating the attenuation of the retroactivity by the PBI circuit. These results suggest that the insulator does help attenuate the retroactivity in biological circuits in TX-TL platform.

Besides, we also investigated the temperature sensitivity of this circuit in TX-TL. We performed the same experiments at 29°C, 33°C, and 37°C, respectively (Supplementary Figure S2). At 29°C, the PBI circuit was able to limit retroactivity to about 20%; however, as temperature increased, the efficiency of insulation decreased to about 40% and 50%; while the control circuit had the same signal reduction among all three temperatures. The results suggested that the PBI circuit is sensitive to reaction temperature in TX-TL and its performance is affected by the temperature. Previous TX-TL characterization experiments suggested that relatively higher temperature would accelerate the molecular reactions involved in transcription and translation, resulting in faster GFP production rate [37]. However, because of the

resource limitation in TX-TL, there could be more intensive competition on resources at higher temperature. As a result, the downstream load might have a better chance to sequester the NRI proteins away from the reporter at higher temperature, which would lead to less attenuation of the loading effect.

2.4 Estimation of Constitutively Expressed Protein Concentrations

In this section, our goal is to derive a simplified model that can be uniquely parameterized from a set of characterization experiments in the bimolecular breadboard system [20, 21]. We base our model on the general phosphorylation-based insulator model posed in [43], but adapt our notation and augment input variables that are present in the biomolecular breadboard system. Because it is an *in vitro* system, the total DNA and inducer concentration in solution are adjustable experimental variables or variables that can be modeled as inputs. It is the freedom of these inputs that allows us to perform experiments and collect data that parameterizes the model.

We begin by introducing a chemical reaction model for the system:

Equation 1

$$NRI + K \xrightarrow{k_{ph}} NRI^P + K,$$

$$NRI^P + Ph \xrightarrow{k_{deph}} NRI + Ph,$$

$$NRI^P \xrightarrow{k_{auto}} NRI,$$

$$p_K \xrightarrow{k_{TX,K}} m_K \xrightarrow{k_{TL,K}} K^u \xrightarrow{k_{f,K}} K,$$

$$p_{Ph} \xrightarrow{k_{TX,Ph}} m_{Ph} \xrightarrow{k_{TL,Ph}} Ph^u \xrightarrow{k_{f,Ph}} Ph,$$

$$p_{NRI} \xrightarrow{k_{TX,N}} m_{NRI} \xrightarrow{k_{TL,N}} NRI^u \xrightarrow{k_{f,N}} NRI,$$

$$m_K \xrightarrow{\delta_m} \emptyset,\ m_{Ph} \xrightarrow{\delta_m} \emptyset,\ m_{NRI} \xrightarrow{\delta_m} \emptyset,$$

$$NRI^P + p_{GlnA} \underset{k_u}{\overset{k_b}{\leftrightarrows}} NRI^P{:}p_{GlnA} \xrightarrow{k_{cat}} NRI^P + p_{GlnA} + GFP,$$

$$NRI^P + p_{GlnA,L} \underset{k_u}{\overset{k_b}{\leftrightarrows}} NRI^P{:}p_{GlnA,L} \xrightarrow{k_{cat}} NRI^P + p_{GlnA,L} + RFP,$$

where $K, P, NRI,$ and NRI^P denote the kinase, phosphatase, unphosphorylated NRI, and phosphorylated NRI protein, pGlnA is the GlnA promoter, pGlnA,L is the GlnA promoter encoding for other competing genes, NRI^u is the unfolded form of NRI protein, K^u is the unfolded form of kinase, Ph^u is the unfolded form of phosphatase, and $\emptyset$ represents a macro state of all degraded mRNA. We also use the notation X^{tot} when needed to denote the total amount of protein X, where $X = K, Ph, NRI$. This notation will be convenient for our analysis in the sequel.

Since p_X represents a constitutive promoter for $X = K, Ph, NRI$, the total kinase, phosphatase, and NRI protein are produced constitutively. An assay with GFP shows that without additional proteases added into the bimolecular breadboard system, protein degradation is negligible [22]. Thus, we can approximate the total amount of NRI protein at a particular time t expressed under the pCon promoter

using GFP expression expressed under the same promoter and ribosome binding site (RBS) as a proxy. This total amount of NRI, we will denote as NRI^{tot}.

We also note that an alternative approach to estimate $NRI^{tot}(t)$ is to assay the expression of a NRI-GFP fusion protein. However, this approach may significantly alter the phosphorylation dynamics of the NRI protein, since it acts as a substrate for the kinase. Therefore, we will express GFP separately on the pCon promoter with the same RBS and use it to estimate concentration from arbitrary units of fluorescence.

Because there are differences in the transcription and translation rates, and folding of GFP and NRI, we do not expect the estimated concentration of GFP at time t will be identical to the concentration of the NRI protein at time t. We can account for these differences dynamically in a mass action model of NRI and GFP dynamics. If we consider NRI constitutive expression in a simple isolated system with no kinase or phosphatase activity NRI^{tot}, e.g. with the chemical reaction system

$$p_{Con} \xrightarrow{k_{TX,N}} m_{NRI} \xrightarrow{k_{TL,N}} NRI^{u} \xrightarrow{k_{f,N}} NRI,$$

we see that

$$\frac{dNRI}{dt} = k_{f,N} NRI^{u},$$

$$\frac{dNRI^{u}}{dt} = k_{TL,N} m_{NRI},$$

$$\frac{dm_{NRI}}{dt} = k_{TX,N} p_{Con} - \delta_{m} m_{NRI}.$$

The total NRI protein at time t is ultimately a function of $m_{NRI}(t)$. Since the dynamics of m_{NRI} can be viewed as a scalar linear system with static step input pCon, we can solve analytically for $NRI^{tot}(t)$ to obtain:

Equation 2

$$
\begin{aligned}
NRI(t) &= NRI(t_0) + k_{f,N}\int_0^t NRI^u d\tau \\
&= NRI(t_0) + k_{f,N}\int_0^t \left[NRI^u(t_0) + \int_0^\tau \left(k_{TL,N} m_{NRI}\right) d\xi\right] d\tau \\
&= NRI(t_0) \\
&\quad + k_{f,N}\int_0^t \Bigg[NRI^u(t_0) \\
&\quad + \int_0^\tau \left(k_{TL,N} m_{NRI}(t_0) + k_{TL,N} k_{TX,N} p_{Con}\left(\frac{1 - e^{-\delta_m \xi}}{\delta_m}\right)\right) d\xi\Bigg] d\tau \\
&= \frac{k_{f,N} k_{TX,N} k_{TL,N}}{\delta_m} p_{Con}\left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{-\delta_m t}\right),
\end{aligned}
$$

whenever $NRI(t_0) = m_{NRI}(t_0) = NRI^u(t_0) = 0$. To reflect the experimental conditions of our system, we have assumed that the initial mRNA, unfolded and folded kinase, phosphatase, and NRI concentrations are zero. Notice that in deriving this expression, we have made no assumption about time-scale separation. While such arguments are valid since the folding dynamics proceed at a much slower rate than the transcription and translation dynamics, they are unnecessary for estimating NRI at time t. Finally, it is worth noting that we assume the transcription and

translation reactions proceed as first order reactions, which is valid as long as our DNA concentrations (typically in the nM range) are much less than the concentrations of RNA polymerases, ribosomes, chaperone proteins, etc. (typically in the μM range [20]).

It is worth noting that model for the mRNA species m_{NRI} is qualitatively consistent with our experimental studies of mSpinach expression in the transcription-translation system. To demonstrate this, we consider a model of the same functional form as *equation 2*, but with a constitutive promoter and coding sequence of the same length as the mSpinach transcript. This yields

$$m_S(t) = \frac{k_{TX} p_{Con}}{\delta_m}(1 - e^{-\delta_m t}),$$

where

$$k_{TX} = \frac{k_{r,bp}}{L(m_S)} k_{isom}$$

is estimated with $k_{r,bp} = 60\ bp/s$ (the approximate mean of a variety of media-dependent rates found in [54]), $L(m_S) = 98\ bp/nM$ is the length of mSpinach aptamer without a tRNA scaffold [55], and $k_{isom} = 6.3{\times}10^{-2}\ s^{-1}$ is the forward rate of open complex formation from the closed complex.

From this, it is possible to estimate the rate of mRNA degradation,

$$\delta_m = \frac{k_{TX} p_{Con}}{m_S(t=120)} = 3{\times}10^{-2} s^{-1},$$

where $m_S(t = 120)$ is the expression of mSpinach at time $t = 120$ minutes and is an approximation of m_S steady state expression if the system were to continue to run indefinitely. The time point $t = 120$ minutes, is critical to consider for our biomolecular breadboard system. Previous results generated with the TX-TL system [22, 56] suggest that protein production rates typically maintain as a constant within 120 minutes of the reactions. After 120 minutes, we can see a decrease in mRNA concentration accordingly (Figure 3A) as well as slower protein production (Figure 1B). Thus, an empirical upper bound on time horizon for our model is approximately 120 minutes after the reaction is initiated.

It is also important to mention that with the exception of the mRNA species m_{NRI} of NRI protein in our model, the species associated with NRI do not settle at a stable steady state. This aspect of our model is consistent with the behavior of biocircuit expression for an initial window of time in the biomolecular breadboard system. In this *in vitro* system, the auxiliary proteins NRI, K, Ph, and even GFP (expressed by pGlnA) do not achieve a steady state in the traditional manner (due to detailed balance of production, and a combination of degradation and dilution effects). Rather, they continue to increase in concentration until all transcriptional and translational resources are exhausted. Thus, because we are interested in the dynamic behavior of the phosphorylation-based insulator and drawing comparisons of its *in vitro* behavior to *in vivo* behavior, we will focus our subsequent modeling efforts in the time window $t \in (0,120)$ minutes where fuel, energy, and other transcriptional and

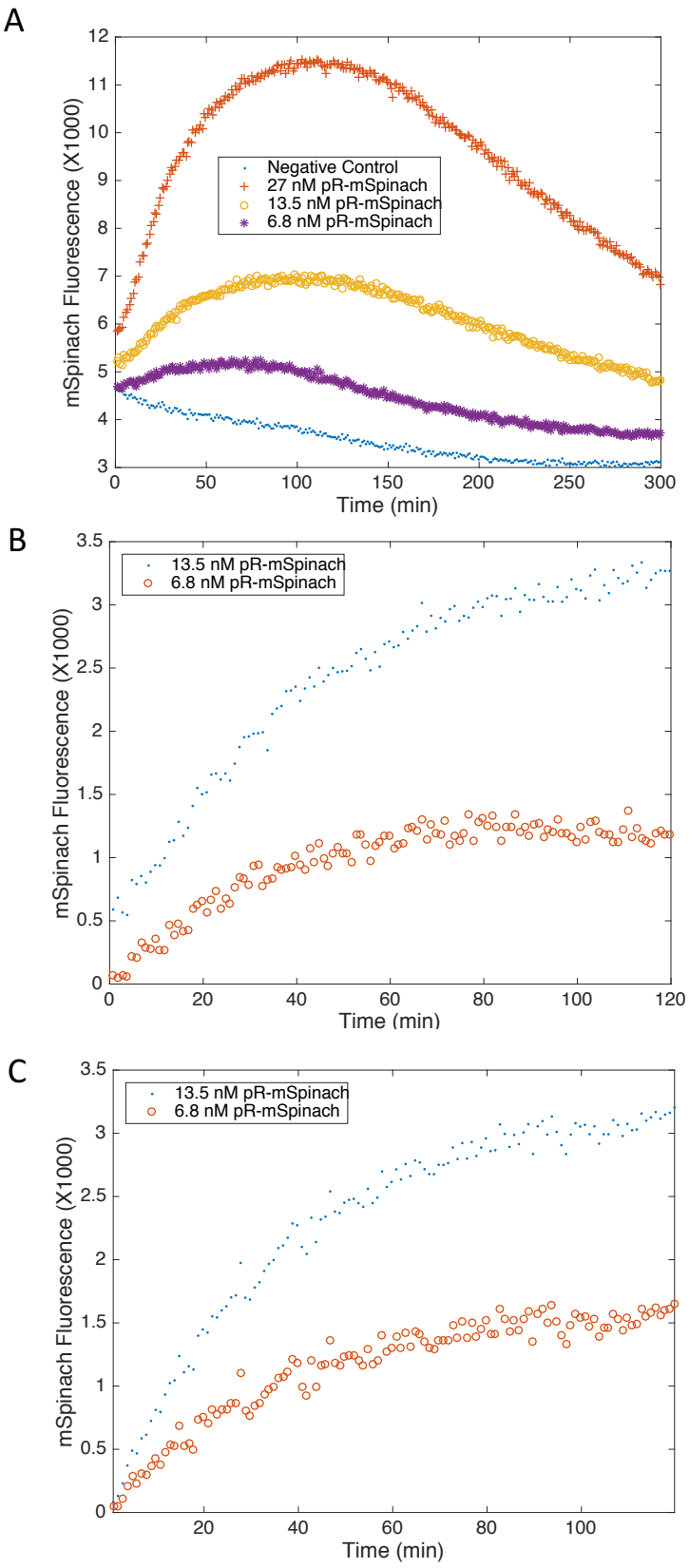

A
mSpinach Fluorescence (X1000)
Time (min)
Negative Control
27 nM pR-mSpinach
13.5 nM pR-mSpinach
6.8 nM pR-mSpinach
B
mSpinach Fluorescence (X1000)
Time (min)
13.5 nM pR-mSpinach
6.8 nM pR-mSpinach
C
mSpinach Fluorescence (X1000)
Time (min)
13.5 nM pR-mSpinach
6.8 nM pR-mSpinach

Figure 3 **A**: Data featuring mSpinach expression on linear DNA with 100 bp of protection. The transcriptional unit consists of an OR1-OR2-pR promoter, followed by the mSpinach (no scaffold) RNA aptamer coding sequence and the T500 terminator. Arbitrary fluorescence units of mSpinach expression is plotted against time. Subtracting the background, we see that mSpinach expression nearly doubles as DNA concentration doubles. Past t = 120 minutes, mSpinach expression decreases, presumably because linear DNA template has degraded or transcriptional resources are exhausted. Our time horizon of interest for the model will thus be in the interval of $t \in (0, 120)$. **B**: Data featuring mSpinach expression driven by the OR1-OR2-pR promoter at 13.5 nM and 6.8 nM concentration from the time interval of 0 to 120 minutes. mSpinach expression dynamics in the time horizon of interest feature a phase of steep linear growth and then saturation towards an asymptotic limit. **C**: A simulation of mSpinach expression, driven by a constitutive promoter at 6.8 nM and 13.5 nM DNA concentration. Notice that the model is able to capture the qualitative effects of mSpinach expression.

translational resources are still abundant. In doing so, we do not preclude the possibility of genes competing with each other for the finite resources available in the *in vitro* system. Our time frame of interest is thus when transcriptional and translational machinery is available and functional, but in finite supply (mimicking *in vivo* conditions).

Using the parameters that we have calculated, we plot the outcome of a simulation against expression data for mSpinach in Figure 3. The output is simulated with additive white noise, replicating the measurement noise present in the plate reader (refer to the trajectory of the negative control). We use a biocircuit expressing mSpinach with the constitutive promoter (pOR1-OR2 from the λ regulatory operon). Notice that the functional form of $m_{NRI}(t)$ adequately describes the qualitative behavior of mRNA expression in the breadboard system until $t \cong 120$ minutes. The rate at which mSpinach saturates is determined by the δ_m parameter and its steady

state value is given as $k_{TX}p_{Con}/\delta_m$. These experiments with the mSpinach RNA aptamer show that our model, while simple in its formulation, is sufficiently complex to describe transcriptional dynamics in the transcription-translation system for the first two hours. Thus, we will not attempt to model expression when the transcription-translation system depletes it resources; at this point gene expression is strongly competitive, and production and degradation rates are largely determined by the available ATP, rNTPs, amino acids, etc. in the system.

We also know that the folding rates of the GFP protein are different from those of NRI protein. Thus, to estimate the ratio in folding rates, we use the K-fold protein folding simulation software developed in [57]. Based on previous studies [58], we use the following equations to estimate the mRNA transcription rate and the protein translation rate:

$$\frac{dm_X}{dt} = \alpha - \delta m_X, \qquad \frac{dX}{dt} = \kappa m_X - \gamma X,$$

where m_X is the concentration of mRNA for protein X ($X = NRI, GFP$), α is the rate of production of the mRNA for protein X, δ the rate of degradation of the mRNA, κ is the rate of translation of mRNA and γ is the rate of degradation of protein X. The value of α increases with the strength of the promoter while the value of κ increases with the strength of the RBS [58]. Because we are using the same promoter and RBS for both NRI and GFP genes, they should share the same mRNA transcription rate and protein translation rate, despite their differences in gene length in first approximation.

We express the rates of transcription, translation, and folding for NRI in terms of GFP rates of transcription, translation, and folding (respectively) as follows:

$$k_{TX,N} = k_{TX,G} \equiv \alpha_{TX} k_{TX,G},$$

$$k_{TL,N} = k_{TL,G} \equiv \alpha_{TL} k_{TL,G},$$

$$k_{f,N} = \frac{0.85 s^{-1}}{1.23 s^{-1}} k_{f,G} \equiv \alpha_f k_{f,G},$$

where "≡" denotes a definition of α_j. Our model for GFP expression under the pCon promoter is similarly expressed as

$$p_{Con} \xrightarrow{k_{TX,G}} m_{GFP} \xrightarrow{k_{TL,G}} GFP^u \xrightarrow{k_{f,G}} GFP^{tot}.$$

The model derived for $GFP(t)$ follows an analogous derivation as the model for $NRI^{tot}(t)$. Thus, using *equation 2*, it is straightforward to show that $NRI^{tot}(t)$ concentration can be expressed as

Equation 3

$$NRI^{tot}(t) = \frac{\alpha_{TX} k_{TX,G} \alpha_{TL} k_{TL,G} \alpha_f k_{f,G} p_{Con}}{\delta_m} \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{\delta_m t} \right)$$

$$= \alpha_{TX} \alpha_{TL} \alpha_f GFP(t).$$

We see that by scaling the GFP concentration by the appropriate ratios at time t, we can obtain an estimate for NRI^{tot}. The above formula holds as long as the concentration of pCon promoter expressing GFP is the same as the concentration of pCon promoter expressing NRI protein with the same RBS. Otherwise, a ratio to

account for the scaling between the two should also be incorporated into the above relation.

To summarize, we have posed a basic model for constitutive expression of NRI protein; the model has a closed form analytical expression that allows estimation of total NRI protein as a function of time. Our model relies on a basic set of chemical reactions describing the processes of transcription and translation. To justify our model at the transcriptional level, we have performed an experimental assay using the mSpinach RNA aptamer to ascertain the dynamics of mRNA expression in the biomolecular breadboard system. Our simulations and experimental data appear to match for up to the first two hours of the experiment, based on parameters extracted from various references, suggesting that our model is accurate in a time horizon of interest. We thus restrict our attention to this time horizon, as it represents the horizon in which transcription and mRNA degradation proceed unperturbed. Further, evidence in [37] suggests that ribosomal activity proceeds unhindered in the first two hours.

We also observe that an analogous line of reasoning can be applied to estimating Ph^{tot} and K^{tot}. We do not repeat the derivation here, as it only requires a change in notation. However, we emphasize that because of these observations, in the sequel we will refer to $Ph^{tot}(t)$, $K^{tot}(t)$, and $NRI^{tot}(t)$ as additional input variables (so long as we are modeling the appropriate time horizon). Additionally, it is the ratio of Ph^{tot} and K^{tot} that matter as a functional input in the system identification process and not the individual concentrations that matter. Further, it is by levering the inputs

Ph^{tot}/K^{tot}, and NRI^{tot} that we are able to identify the parameters of a simplified model uniquely.

2.5 Derivation of A Simplified Model for the PBI

In this section, we derive a simplified model of the phosphorylation-based insulator using the chemical reaction system (*equation 1*). Examining the full chemical reaction system (*equation 1*), we obtain the following state space model from the law of mass action:

Equation 4

$$\begin{aligned}
\frac{dNRI^P}{dt} &= k_{ph}NRI \cdot K - k_{deph}NRI^P Ph \\
&\quad -k_b NRI^P(p_{GlnA} + p_{GlnALoad}) - k_{auto}NRI^P \\
&\quad +(k_u + k_{cat})(NRI^P{:}p_{GlnA} + NRI^P{:}p_{GlnALoad}),
\end{aligned}$$

$$\frac{dNRI}{dt} = k_{deph}NRI^P Ph - k_{ph}NRI \cdot K + k_{f,N}NRI^u + k_{auto}NRI^P,$$

$$\frac{dNRI^u}{dt} = k_{TL,N}m_{NRI},$$

$$\frac{dm_{NRI}}{dt} = k_{TX,N}p_{NRI} - \delta_m m_{NRI},$$

$$\frac{dK}{dt} = k_f K^u,$$

$$\frac{dK^u}{dt} = k_{TL,K}m_K,$$

$$\frac{dm_K}{dt} = k_{TX,K}p_K - \delta_m m_K,$$

$$\frac{dPh}{dt} = k_f Ph^u,$$

$$\frac{dPh^u}{dt} = k_{TL,Ph} m_{Ph},$$

$$\frac{dm_{Ph}}{dt} = k_{TX,Ph} p_{Ph} - \delta_m m_{Ph},$$

$$\frac{dNRI^P{:}p_{GlnA}}{dt} = k_b NRI^P p_{GlnA} - (k_u + k_{cat})(NRI^P{:}p_{GlnA}),$$

$$\frac{dNRI^P{:}p_{GlnALoad}}{dt} = k_b NRI^P p_{GlnALoad} - (k_u + k_{cat})(NRI^P{:}p_{GlnALoad}),$$

$$\frac{dGFP}{dt} = k_{cat} NRI^P{:}p_{GlnA}, \qquad \frac{dRFP}{dt} = k_{cat} NRI^P{:}p_{GlnALoad}.$$

The dimension of the state-space model is 14 and because of the presence of bimolecular reactions, it is nonlinear in the state of the system. Thus, it is difficult to obtain a closed form expression for the solution to the system. However, we will systematically impose a series of modeling assumptions that are physiologically plausible, but which greatly reduce the complexity of the model.

First, notice that the total concentration of K, Ph, and NRI, denoted as K^{tot}, Ph^{tot}, and NRI^{tot}, depends only on the transcription and translation reactions. Thus, if we consider the transcription and translation dynamics of K^{tot}, Ph^{tot}, and

$$\begin{aligned} NRI^{tot} &= NRI^P + NRI + NRI^P{:}p_{GlnA} + NRI^P{:}p_{GlnALoad} \\ &= k_{f,N} NRI^u, \end{aligned}$$

in isolation, we can use the results of the previous section to obtain a closed form expression for their total concentration as follows:

$$NRI^{tot}(t) = \frac{k_{f,N}k_{TX,N}k_{TL,N}}{\delta_m}p_{NRI}\left(\frac{t^2}{2} - \frac{1}{\delta_m^2}e^{-\delta_m t}\right),$$

$$K^{tot}(t) = \frac{k_{f,K}k_{TX,K}k_{TL,K}}{\delta_m}p_{K}\left(\frac{t^2}{2} - \frac{1}{\delta_m^2}e^{-\delta_m t}\right),$$

$$Ph^{tot}(t) = \frac{k_{f,Ph}k_{TX,Ph}k_{TL,Ph}}{\delta_m}p_{Ph}\left(\frac{t^2}{2} - \frac{1}{\delta_m^2}e^{-\delta_m t}\right).$$

These total concentrations can be viewed as time varying parameters. If we had a way of quantifying the rate of transcription, translation, and folding of the individual proteins in the transcription-translation system, we could predictively estimate the trajectories of NRI^{tot}, K^{tot}, and Ph^{tot} over time. However, we do not have these parameters, and thus it is advantageous to employ the previous section's approach. With similar arguments, we can argue that the total concentration of these proteins can be expressed as the total concentration of a reporter molecule multiplied by a scaling constant (see *equation 3*). Thus, using a separate assay to quantify constitutive expression of a reporter molecule under a given constitutive promoter (and a calibration curve to convert fluorescence to molar concentration), we can use the reporter molecule as a proxy for estimating the true molar concentration of NRI, kinase or phosphatase. Therefore, we can avoid the problem of estimating transcriptional, translational, and folding rates of heterogeneous proteins while obtaining an estimate of the functional protein concentrations. Moreover, the result holds for all t in which RNA expression increases linearly ($t \leq 120$).

We formalize this assumption as follows:

Assumption 1: We suppose that for all $t \in [0,120]$, $K^{tot}(t)$, $Ph^{tot}(t)$, and $NRI^{tot}(t)$ are known parameters. This assumption thus allows us to eliminate the dynamics of folded kinase, unfolded kinase, folded NRI, unfolded NRI, folded phosphatase, unfolded phosphatase, and all mRNA dynamics.

The remaining dynamics of the system are thus given as:

$$\frac{dNRI^P}{dt} = k_{ph}NRI \cdot K - k_{deph}NRI^P Ph$$

$$-k_b NRI^P(p_{GlnA} + p_{GlnALoad}) - k_{auto}NRI^P$$

$$+(k_u + k_{cat})(NRI^P{:}p_{GlnA} + NRI^P{:}p_{GlnALoad}),$$

$$\frac{dNRI}{dt} = k_{deph}NRI^P Ph - k_{ph}NRI \cdot K + k_{f,N}NRI^u + k_{auto}NRI^P,$$

$$\frac{dNRI^P{:}p_{GlnA}}{dt} = k_b NRI^P p_{GlnA} - (k_u + k_{cat})(NRI^P{:}p_{GlnA}),$$

$$\frac{dNRI^P{:}p_{GlnALoad}}{dt} = k_b NRI^P p_{GlnALoad} - (k_u + k_{cat})(NRI^P{:}p_{GlnALoad}),$$

$$\frac{dGFP}{dt} = k_{cat}NRI^P{:}p_{GlnA},$$

$$\frac{dRFP}{dt} = k_{cat}NRI^P{:}p_{GlnALoad}.$$

Next, we assume that the phosphorylation and dephosphorylation reactions occur at a much faster time scale than the production of GFP or RFP and the binding (and unbinding) reactions of NRI^P to DNA to form (or disintegrate) activator-DNA complex. We justify the latter assumption through experimental observations that

observe phosphorylation rates on the order of $10^6\ min^{-1}$. Transcription factor binding rates are less characterized but typically binding and unbinding rates of a transcription factor (e.g. LacI) are $O(10^{-1})\ min^{-1}$ and $O(10)\ min^{-1}$ respectively [59].

We formalize these assumptions as follows:

Assumption 2: We suppose that $k_{ph}, k_{deph} \gg k_u, k_{cat}, k_b, k_{auto}$.

Next, we suppose that the amount of DNA bound NRI^P is smaller than the amount of free NRI^P and unphosphorylated NRI and that total NRI can be approximated as the sum of unbound NRI^P and NRI. Put another way, we assume that the molar concentration of unbound NRI protein is substantially larger than the molar concentration of DNA-bound NRI protein. This will certainly be the case since the pGlnA and pGlnAload DNA concentration will be in the nM range while the protein concentration of NRI will be in the μM range (refer to the arguments in the previous section and Figure 3). From the above reactions and assumptions, we then can write the dynamics of NRI^P using the approximate conservation law

$$NRI^{tot} \cong NRI^P + NRI$$

as follows:

$$\frac{dNRI^P}{dt} = k_{ph}(NRI^{tot} - NRI^P)K^{tot} - k_{deph}Ph^{tot}NRI^P.$$

Since phosphorylation and dephosphorylation occurs at a much faster rate than GFP and RFP production (our ultimate time-scale of interest) and reasonably faster than the binding dynamics of the $\mathrm{NRI}^{\mathrm{P}}$ transcriptional activator, we can solve the fast dynamics of $\mathrm{NRI}^{\mathrm{P}}$ to obtain an analytical expression for the equilibrium point NRI_e^P. At steady state, we have

$$\begin{aligned} 0 &= \frac{dNRI^P}{dt} \\ &= k_{ph}(NRI^{tot} - NRI^P)K^{tot} - k_{deph}Ph^{tot}NRI^P, \end{aligned}$$

which implies

$$\begin{aligned} NRI_e^P &= \frac{k_{ph}NRI^{tot}K^{tot}}{k_{ph}K^{tot} + k_{deph}Ph^{tot}} \\ &= \frac{\frac{k_{ph}}{k_{deph}}NRI^{tot}}{\frac{k_{ph}}{k_{deph}} + \frac{Ph^{tot}}{K^{tot}}} \equiv \theta(NRI^{tot}, K^{tot}, Ph^{tot}), \end{aligned}$$

where "≡" denotes the definition of the function $\theta(NRI^{tot}, K^{tot}, Ph^{tot})$.

In this assumption, we also assume that k_{auto} is negligible compared to other rates. This is a reasonable assumption since spontaneous dephosphorylation proceeds at a slow rate — the ΔG of spontaneous dephosphorylation is very large [60].

The final assumption we leverage is that the rates of GFP and RFP production, relative to the binding dynamics of $\mathrm{NRI}^{\mathrm{P}}$ are much slower. Specifically, we suppose that:

Assumption 3: $k_b, k_u \gg k_{cat}$.

This assumption can be justified, since the production of a folded protein such as GFP takes at least ten to fifteen minutes [61] while the binding and unbinding rates are typically on the order of hundredths of seconds and seconds, respectively [59]. Thus, we can solve for the steady state of the DNA-activator complexes $\mathrm{NRI^P}$: pGlnA and $\mathrm{NRI^P}$: pGlnALoad. The result is analogous to the classical Michaelis–Menten model, with $V_{max} = NRI_e^P$ and $K_M = (k_u + k_{cat})/k_b$. We omit the derivation, as it follows the standard derivation for a two-substrate one-enzyme model:

Equation 5

$$\frac{dGFP}{dt} = k_{cat} \frac{\frac{\theta p_{GlnA}^{tot}}{K_M}}{1 + \frac{p_{GlnA}^{tot} + p_{GlnALoad}^{tot}}{K_M}},$$

$$\frac{dRFP}{dt} = k_{cat} \frac{\frac{\theta p_{GlnALoad}^{tot}}{K_M}}{1 + \frac{p_{GlnA}^{tot} + p_{GlnALoad}^{tot}}{K_M}},$$

where θ denotes $\theta(NRI^{tot}, K^{tot}, Ph^{tot})$. This completes the derivation of our simplified model. In the next section, we will analyze the structure of the model, determine which of the parameters are globally identifiable, and under what circumstances identifiability holds.

2.6 System Identification of the Simplified PBI Model

A. Theoretical Analysis

In the derivation of our model we have made a point to retain the experimental parameters NRI^{tot}, K^{tot}, Ph^{tot}, p_{GlnA}^{tot} and $p_{GlnALoad}^{tot}$. These parameters can be viewed as experimentally controllable, in that we can directly control the DNA concentration of promoters p_{GlnA}^{tot} and $p_{GlnALoad}^{tot}$. Additionally, by adjusting the underlying constitutive promoters driving the expression of NRI, K, and Ph we can effectively tune the quantities NRI^{tot}, K^{tot}, and Ph^{tot}. We note that this type of control over the concentration of DNA as well as total protein concentration is not typically achievable *in vivo*, unless inducers are employed or different replication origins are cloned into a plasmid (which introduces variability in copy number from cell-to-cell). However, this advantage in the biomolecular breadboard is precisely the capability required to explore the problem of parameter estimation and determine if our simplified model is globally identifiable.

Since our calibration curves allow us to estimate GFP concentration from arbitrary fluorescence units, we will focus our attention on the GFP dynamics. Furthermore, notice that the forms of dynamics of both reporter molecules are identical. Thus, it suffices to analyze the identifiability of parameters with respect to the output dynamics of the GFP reporter molecule, since it will yield the same result as studying identifiability with respect to RFP output dynamics. Recalling our assumptions from the previous section, we will also make a point to study the behavior of the system within the time horizon of interest captured by our model, $t \in [0, \tau_{max})$, where τ_{max} is the initiation of the resource depletion phase in our transcription-translation system.

Our goal is to determine whether this model is globally structurally identifiable with respect to the parameters K_M, k_{cat}, k_{ph}, and k_{deph}, given the inputs NRI^{tot}, K^{tot}, Ph^{tot}, p_{GlnA}^{tot}, and $p_{GlnALoad}^{tot}$. Notice that the inputs do not enter the dynamics of the system in a linear fashion. Indeed, the simplified system (*equation 5*) is of the form: $dx/dt = f(U, \Theta)$, where f is nonlinear with respect to U and Θ, in which $U = (NRI^{tot}, Ph^{tot}, K^{tot}, p_{GlnA}^{tot}, p_{GlnALoad}^{tot})$ and $\Theta = (K_M, k_{cat}, k_{ph}, k_{deph})$.

Furthermore,

$$f(U, \Theta) = f_1(U_1, \Theta_1) f_2(U_2, \Theta_2),$$

where

$$U_1 = (NRI^{tot}, Ph^{tot}, K^{tot}), \qquad \Theta_1 = \left(k_{ph}, k_{deph}\right),$$

$$U_2 = (p_{GlnA}^{tot}, p_{GlnALoad}^{tot}), \qquad \Theta_2 = (K_M, k_{cat}).$$

Notice that $f_1 = NRI_e^P = \theta(NRI^{tot}, Ph^{tot}, K^{tot})$ takes the form of a Hill function with Ph^{tot}/K^{tot} as its substrate and

$$f_2 = k_{cat} \frac{\frac{p_{GlnA}^{tot}}{K_M}}{1 + \frac{p_{GlnA}^{tot} + p_{GlnALoad}^{tot}}{K_M}}.$$

This multiplicative decomposition provides a key insight: our system dynamics is the product of two Hill functions with distinct inputs for each Hill function. This suggests that from a system identification standpoint, we can attempt a series of

experiments that perturb one of the Hill functions while holding the other constant and vice versa to tease out the parameters for each.

To obtain insight into the what parameters in the Hill functions are identifiable, we invert the system dynamics to obtain

$$\frac{1}{dGFP/dt} = \frac{1}{f_1 f_2} = \frac{(p_{GlnA}^{tot} + p_{GlnALoad}^{tot}) + K_M}{\theta k_{cat} p_{GlnA}^{tot}},$$

which indicates that the parameter identification is a linear regression problem, and after some arrangement we find that

$$\frac{\frac{dGFP}{dt}(p_{GlnA}^{tot} + p_{GlnALoad}^{tot})}{p_{GlnA}^{tot}} = f_1 k_{cat} - K_M \frac{dGFP/dt}{p_{GlnA}^{tot}}.$$

Thus, when the experimental input $p_{GlnALoad}^{tot}$ is set to 0 nM, we obtain a linear regression problem in estimating slope K_M and intercept $k_{cat} f_1$. Further, if we enforce that $Ph^{tot} = 0$, then f_1 reduces to NRI^{tot}, a known input value that completes the decomposition. Thus, by enforcing these two input constraints, we obtain a linear regression problem that effectively estimates k_{cat} and K_M. By varying the total DNA concentration p_{GlnA}^{tot} we can thus vary the rate of change of GFP, $dGFP/dt$, and obtain data to optimize k_{cat} and K_M. Once k_{cat} and K_M are estimated, we can then use a similar line of arguments to back out an estimate for the ratio k_{ph}/k_{deph}.

In particular, we consider a nominal operating concentration of p_{GlnA}^{tot} and $p_{GlnALoad}^{tot}$, and write $\gamma = (U_2, \Theta_2)/f_2$ and $k_\gamma = k_{ph}/k_{deph}$, then taking the reciprocal of $dGFP/dt$ we obtain

$$\frac{1}{dGFP/dt} = \frac{1}{\gamma}\left(\frac{k_\gamma + \frac{Ph^{tot}}{K^{tot}}}{k_\gamma NRI^{tot}}\right)$$

and define $\tilde{Y} = \gamma NRI^{tot}/(\frac{dGFP}{dt})$ we see that

$$\tilde{Y} = 1 + \frac{1}{k_\gamma}\frac{Ph^{tot}}{K^{tot}}.$$

Therefore, by transforming the problem into the reciprocal space, we see that $k_\gamma = k_{ph}/k_{deph}$ is a uniquely identifiable parameter. That is, the problem of estimating k_γ can be expressed as a linear regression problem with k_γ as the reciprocal of the slope and an intercept of unity. The fact that we can write the parameter estimation problem for (k_{cat}, K_M, k_{ph}/k_{deph}) as a solution to a system of linear equations thus shows that the model is globally structurally identifiable with respect to (k_{cat}, K_M, k_{ph}/k_{deph}).

In summary, we have derived a simplified model for the phosphorylation-based insulator and shown it is globally identifiable with respect to the output trajectory of GFP. We have shown that in the theoretical scenario where a continuous trajectory of GFP can be obtained to estimate its derivative $dGFP/dt$, the parameters k_{cat}, K_M, and $k_\gamma = k_{ph}/k_{deph}$ can be estimated. These parameters are only estimated through

a series of carefully designed experiments in which specific TX-TL controllable experimental variables are tuned. In the next section, we discuss the results of these experiments and numerical estimation of this data from time-series data.

B. Experimental Analysis: Systematic Perturbations of the Phosphorylation-Based Insulator for System Identification

To identify the parameters k_{cat} , K_M , and k_γ, we needed to perturb the phosphorylation-based insulator with the experimental variables designated in our model. In particular, we first needed to perturb the amount of pGlnA promoter producing GFP in the absence of phosphatase Ph^{tot} or NRIIH139N protein. Varying the amount of pGlnA promoter in the system in the absence of phosphatase would enable the estimation of k_{cat} and K_M.

Intuitively, k_{cat} and K_M characterize the enzyme-substrate relationship that the activator protein NRI^P has with the pGlnA promoter — coincidentally, to reveal these parameters we need to eliminate any negative feedback imposed on the activator protein by NRIIH139N phosphatase and vary the substrate concentration p_{GlnA}^{tot} to reveal the kinetic parameters.

Accordingly, we ran a set of TX-TL reactions in which the DNA concentration of pCon promoter driving NRIIH139N expression was 0 nM. We varied the concentration of pGlnA promoter from 0 to 57 nM, expressed on plasmid. From the time series data, we extracted the first thirty minutes of expression dynamics — this time horizon constituted the time frame when amino acids, CoA, NADH, ATP, etc.

were far away from the stage of complete depletion in the TX-TL system. In this time horizon of interest, the expression of GFP is linear with respect to time; therefore, the derivative of GFP is constant and can be fitted using the slope of a linear regression. The results of our linear regression are plotted against the time series data of the experiment in Supplementary Figure S3. The estimates of the $dGFP/dt$ in the time horizon of interest at varying concentrations of pGlnA were used to fit the Hill function parameters k_{cat} and K_M:

$$k_{cat} = 2.33\times10^{-2}\, min^{-1},\ K_M = 1.58\, nM.$$

We emphasize that the key to estimating k_{cat} and K_M is the additional freedom afforded by a control input p_{GlnA}^{tot} in perturbing the system.

The final parameter to estimate was $k_\gamma = k_{ph}/k_{deph}$. In order to estimate k_γ, we needed to fix the pGlnA concentrations, i.e. the concentrations driving expression in the downstream module, and perturb the phosphorylation-based insulator. Specifically, we varied the ratio of kinase (NRIIL16R) to phosphatase (NRIIH139N) in the system by varying the ratio of DNA concentrations for the promoters driving their expression. Doing this, we obtained a series of time-lapse curves of GFP expression over a range of Ph^{tot}/K^{tot} values. Again, we extracted estimates for $dGFP/dt$ using a linear regression over the first thirty minutes of gene expression. The resulting estimates for $dGFP/dt$ were then plotted against varying Ph^{tot}/K^{tot} to fit a Hill function (Figure 4). Using standard linear regression techniques, we then obtained the following estimate:

$$k_\gamma = \frac{k_{ph}}{k_{deph}} = 9.81\times10^{-2}.$$

The ratio k_{ph}/k_{deph} characterizes the balance of rates between phosphorylation and dephosphorylation reactions — although we are unable to infer the individual parameters k_{ph} and k_{deph} we are able to conclude that dephosphorylation occurs at roughly an order of magnitude faster than phosphorylation (all other variables equal). Notice this parameter characterizes the intrinsic chemical reaction rates, rather than the flux or mass action rates that are dependent on kinase and phosphatase concentrations. Thus, to tune the phosphorylation-based insulator we can vary the amount of kinase and phosphatase concentrations, bearing in mind that phosphorylation is slower than dephosphorylation in the TX-TL system.

Further, it is consistent with our intuition that only the ratio of k_{ph} and k_{deph} is identifiable and not the individual parameters. Because the individual parameters characterize processes that are much faster than the time-scales of production of our observer molecule GFP and the imaging system in the plate reader, the only information that can be passed onto the observer molecule is the net outcome of NRI protein's phosphorylated state. Phosphorylation and dephosphorylation are processes that compete against each other to increase the amount of NRI^P and NRI concentration in the system, respectively. Thus, by observing the amount of NRI^P in

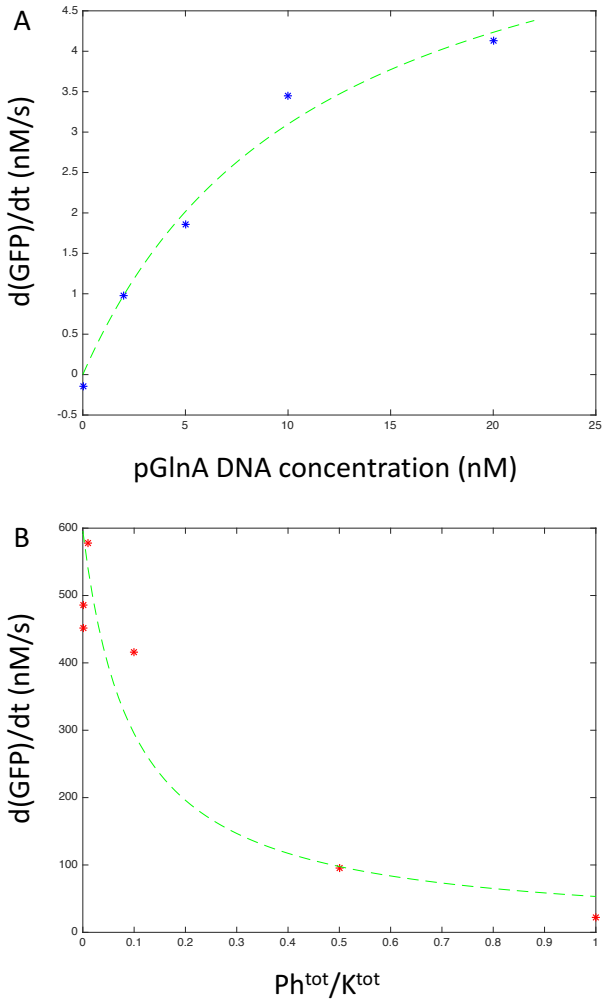


Figure 4 Linear Regression fitting to extract parameters. **A**: A plot of the resulting Hill function $dGFP/dt$ against varying pGlnA. We see the curve follows the form of a Michaelis Menten function which is consistent with our model. **B:** A plot of the resulting Hill function $dGFP/dt$ against varying Ph^{tot}/K^{tot}. Again, the empirical data (starred) matches the functional form of our model.

the system and knowing the concentration of NRI^{tot}, we can deduce the net outcome of the battle, i.e. the ratio k_{ph}/k_{deph}. Notice that without knowledge of NRI^{tot}, we would be unable to estimate k_{ph}/k_{deph}. This again illustrates the importance of having additional experimental inputs for perturbing the system. Even though there

is only one output molecule GFP, we are able to infer three distinct parameters that represent processes from three different time-scales: catalytic synthesis of protein, formation and disassociation of the DNA-activator complex, and phosphorylation/dephosphorylation of NRI protein.

Finally, it is worthwhile to note that the functional form of our model is consistent both quantitatively (small output residual error) and qualitatively. This suggests that our simplified model will serve as a suitable starting point for simulation studies and theoretical analysis.

2.7 Simulations

Having all the parameters identified, we first repeated the results in Figure 2 in simulation. As shown in Figure 5A and 5B, the control circuit suffered from the retroactivity when the load was added. In contrast, the insulator circuit could attenuate the retroactivity significantly. Both the transfer curves and the fold-change comparison are consistent with the experimental results.

Then we wanted to know how would different initial conditions affect the output. We swept different initial conditions for the reporter pGlnA-GFP, the load pGlnA-RFP, and phosphatase over kinase ratio (Ph^{tot}/K^{tot}) (Figure 5C and D). In Figure 5C, we swept [pGlnA-GFP] and [pGlnA-RFPload]. As we can see in the surface plot, more reporter DNA results in more GFP expression. More load DNA brings down the GFP level, which is a result of retroactivity as load DNA competes with reporter DNA for transcription and translation resources, such as RNA

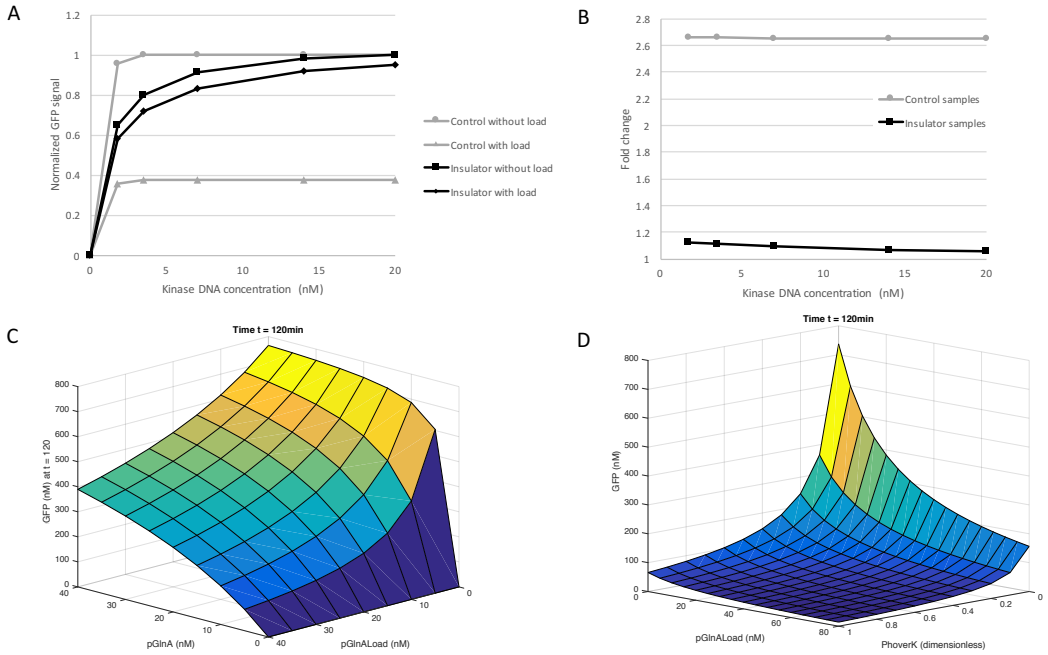


Figure 5 Simulation results of the PBI circuit based on the parameters identified above. **A**: Simulation results of the transfer function curves for controls and insulators with or without load DNA are consistent with the experimental results. **B**: Simulation results of the fold changes of the samples without load over the ones with load are also consistent with the experimental results. **C**: Simulation results of GFP expressions as functions of two DNA inputs. Initial condition: DNA pGlnA-GFP ranging from 0 to 40 nM and DNA pGlnA-RFP (load) ranging from 0 to 40 nM, with 68 nM protein NRI^{tot} and $Ph^{tot}/K^{tot} = 0$. Simulation time is 120 min. D: Simulation results of GFP expressions as functions of DNA load and Ph^{tot}/K^{tot}. Initial condition: DNA pGlnA-RFP (load) ranging from 0 to 80 nM and Ph^{tot}/K^{tot} varying between 0 and 1, with 20 nM DNA pGlnA-GFP and 68nM proteinNRI^{tot}. Simulation time is 120 min.

polymerases and ribosomes. Load DNA or mRNA sequesters those resources away from reporter DNA or mRNA, ending up with less reporter protein made. In Figure 5D, we changed the concentration of the load pGlnA-RFP and the ratio of Ph^{tot}/K^{tot}. As we mentioned above, when certain amounts of phosphatase and kinase are added, the PBI circuit will have a high gain because of phosphorylation and an equally large negative feedback by dephosphorylation. As a result,

retroactivity from downstream load will be attenuated. The simulation results in Figure 5D just showed the exact same idea. By tuning the ratio, we can effectively achieve the same absolute GFP expression level with different absolute load amount. Through this mechanism, the retroactivity is largely attenuated.

2.8 Conclusion

In this work, we investigated the structural identifiability of the phosphorylation-based insulator when implemented in a transcription-translation cell free expression system. We showed that the retroactivity exists in the TX-TL system and the PBI circuit can attenuate the retroactivity significantly. Then we considered a complex model that provided an intricate description of all chemical reactions involved in the PBI circuit. Next, leveraging specific physiologically plausible assumptions, we derived a rigorous simplified model that captures the output dynamics of the phosphorylation-based insulator. We performed standard system identification analysis and determined that the model is globally identifiable with respect to three critical parameters: the catalytic rate associated with the downstream system k_{cat}, an internal parameter in the downstream system characterizing formation of the activator-DNA complex K_M and k_{ph}/k_{deph}, and a ratio describing the intrinsic balance of phosphorylation and dephosphorylation in the PBI circuit. Specifically, we showed that these three parameters were identifiable only when the system was subjected to specific perturbations. We performed these experiments and estimated the parameters. Our experimental results suggest that the functional form of our simplified model is sufficient to describe reporter dynamics and enable parameter

estimation. Besides, our simulations results based on the parameters estimated using above methods confirmed our conclusions from experimental data and previous theoretical predictions. These *in silico* results also showed the power of computational biology and its future applications in guiding biological experiments and synthetic biocircuits design. In general, this research illustrates the utility of the TX-TL cell free expression system as a platform for system identification, as it provides extra control inputs for parameter estimation that typically are unavailable *in vivo*. Future work will investigate the theoretical utility of the TX-TL system as a platform for system identification, parameterization of more complex systems, and the robustness and sensitivity of the phosphorylation-based insulator using our derived model.

Materials and Methods

Plasmids and linear DNAs:

DNA and oligonucleotides primers were ordered from Integrated DNA Technologies (IDT, Coralville, Iowa). Plasmids in this study were designed in Geneious 8 (Biomatters, Ltd.) and were made using standard golden gate assembly (GGA) protocols and maintained in a KL740 strain if using an OR2- OR1 promoter (29°C) or a JM109 strain for all other constructs. Plasmids were mini prepped using Qiagen mini prep kit. BsaI-HF (R3535S) enzyme used in GGA was purchase from New England Biolabs (NEB). Linear DNAs were made by PCRing protein expression related sequences out of GGA constructs using Phusion Hot Start Flex 2X Master Mix (M0536L) from NEB. Rapid assembly procedures were based on [22]. Before being used in the cell-free reaction, both plasmids and PCR products underwent an additional PCR purification step using a QiaQuick column (Qiagen), which removed excess salt. Then PCR products were eluted and stored in deionized water at 4°C for short-term storage and −20°C for long-term storage. All the plasmids used in the work can be found on https://www.addgene.org/.

Fluorescent proteins:

Two fluorescent proteins were used in this work, GFP and RFP. The version of GFP used is deGFP, which is also called eGFP-Del6-229 [62]. The RFP used is mCherry [63].

TX-TL reactions and fluorescence measurements:

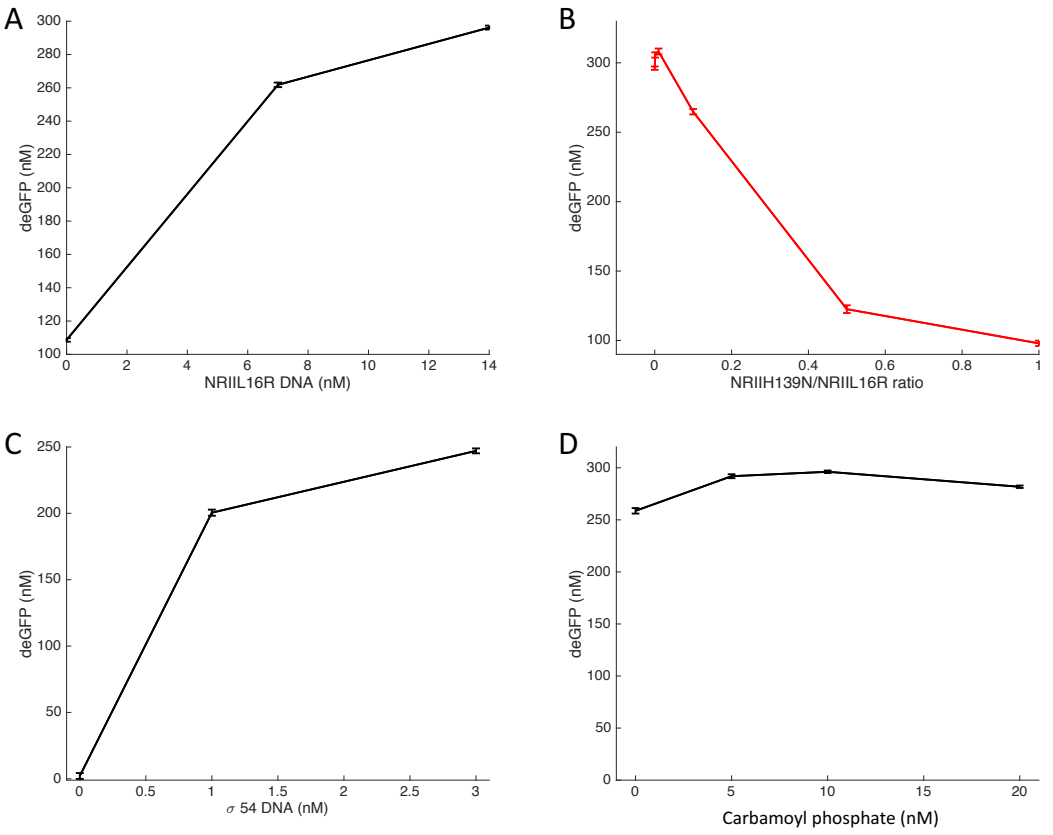
TX-TL reaction mix was prepared and set up according to previous JOVE paper [21]. TX-TL reactions were conducted in a volume of 10 μL in a 384-well plate (Nunc) at 29°C, using a three-tube system: extract, buffer, and DNA. For deGFP, samples were read in a Synergy H1 plate reader (Biotek) using settings for excitation/emission: 485 nm/525 nm, gain 61. All samples were read in the same plate reader, and for deGFP relative fluorescence units (RFUs) were converted to nM of protein using a purified deGFP-His6 standard. Unless otherwise stated, end point measurements are after 2 h of expression at 29°C.

Computational models and simulations:

Data analysis and fitting, model building, and simulations were conducted in MATLAB (R2015b, The MathWorks, Inc.) software.

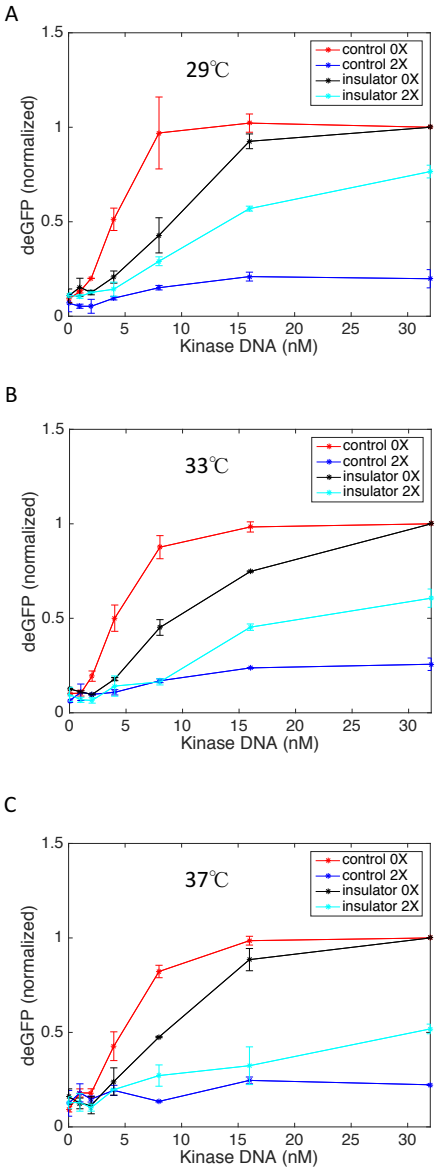
Supplementary Materials

Supplementary Figures

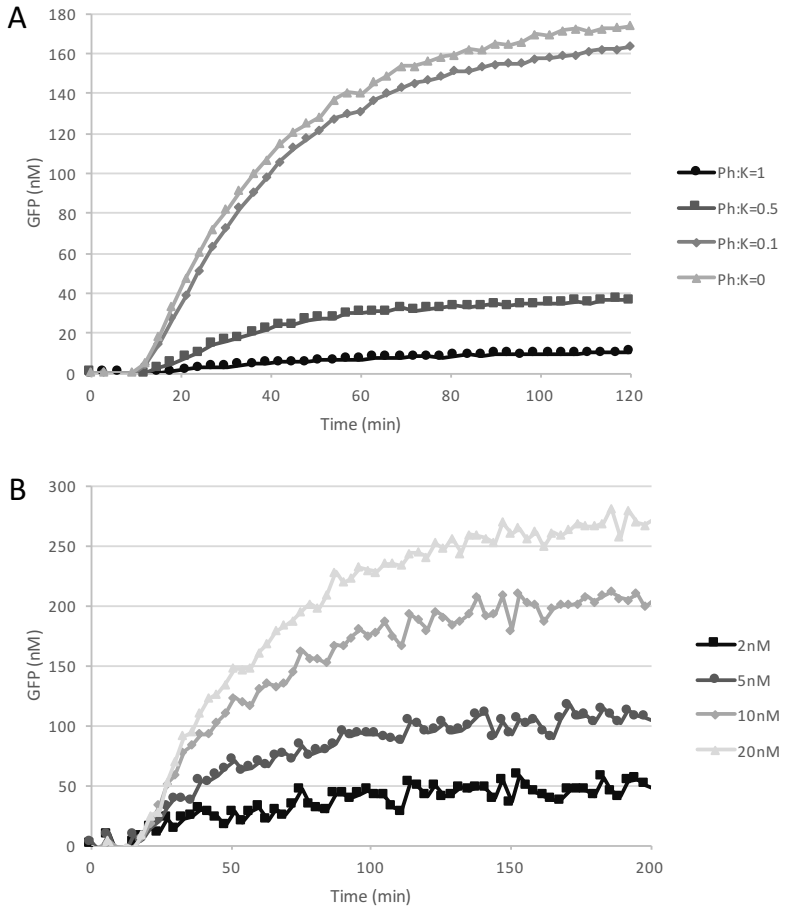


Supplementary Figure S1 Demonstration of the phosphorylation cycle in TX-TL. Before testing the insulation capability of this circuit, we need to demonstrate that phosphorylation and dephosphorylation work in TX-TL. **A:** While keeping the phosphatase at 0 and increasing the concentrations of kinase NRIIL16R linear DNAs, we observed higher pGlnA-deGFP expression as a result of more NRI^P. **B:** When we kept the kinase concentration constant and increased the concentrations of NRIIH139N linear DNAs (higher Ph/K ratio), less pGlnA-deGFP was expressed as a result of dephosphorylation of NRI^P. These results suggested that the phosphorylation cycle worked in TX-TL system. **C:** We also found that σ54 DNA was needed as there was little to no σ54 in the TX-TL cell extract, which is required for the σ54-dependent promoter pGlnA to work. While keeping kinase, phosphatase, and pGlnA-deGFP constant, adding more σ54

DNA significantly increases the expression of deGFP. **D:** We also found that additional phosphate source was recommended but not essential; as shown in the panel, increasing carbamoyl phosphate concentrations did not significantly affect the expression of deGFP. All the deGFP fluorescences have been converted to nM using purified deGFP calibration data.



Supplementary Figure S2 Test the temperature sensitivity of the PBI circuit at three different temperatures, 29°C, 33°C, and 37°C. **A, B, C**: Transfer function curves for controls and insulators with or without load DNA at three different temperatures.



Supplementary Figure S3 GFP expression kinetics were used to estimate the GFP production rate. **A**: Plot of GFP expression while varying the Ph/K ratio. Curves from t = 0 to $\tau_{max} = 30$ min were used to estimate the slope of GFP. **B**: Expression dynamics of GFP for varying amounts of pGlnA with $pLac - Ph = 0\,nM$. These curves enable the estimation of $dGFP/dt$ for $t \leq \tau_{max} = 30\,min$.

Brief introduction on retroactivity and how to attenuate it

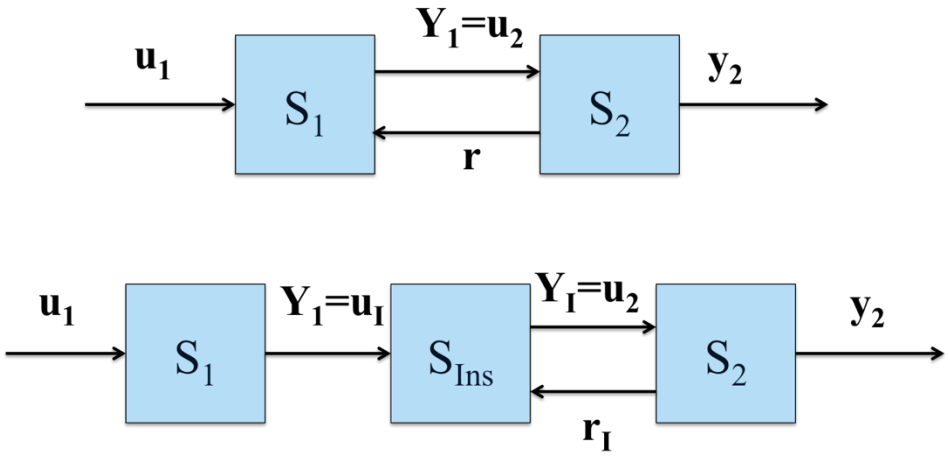
A grand challenge in utilizing modular biological circuits is how to efficiently and robustly connect different modules to yield more complex systems. A difficulty that arises upon the interconnection of two modules in series is a phenomenon called retroactivity [43, 64]. Even if one of the modules is designed to be upstream or driving the input of the second module, a load from the downstream/second module is generated upon interconnection of the two modules that alters the internal dynamics of the first module. There have been significant discoveries on the subject of retroactivity. In [43], the authors developed a rigorous mathematical definition for describing and quantifying retroactivity between two modules. In [65], formal expressions for intramodular and intermodular retroactivity are derived for complex gene transcription networks consisting of nodes, modules, and systems. Additionally, the authors in [66] show that long signal cascades can attenuate retroactivity, while [67] utilizes a time-scale separation strategy to attenuate retroactivity. Finally, in [68] the authors show that attenuating retroactivity can have the unwanted effect of amplifying high frequency noise in gene expression.

An insulator is a biocircuit designed to insulate two biocircuit modules connected in series from the effects of retroactivity. An upstream module S_1 has input u_1 and output y_1 while the downstream module S_2 has input u_2 and output y_2 (Supplementary Figure S3). In a system free of retroactivity effects, the signal $u_2 = y_1$ and there is no signal that maps from S_2 to S_1. However, in biological systems, the output of a system is often a molecule that is consumed or incorporated

in a downstream process. In terms of our model, once S_2 is interconnected to S_1, a retroactive input r from S_2 to S_1 that modifies the dynamics of the output y_1 must be incorporated in our model. The variable r is referred to as retroactivity to the output y_1 and also retroactivity to the input u_2. A general framework is derived in [68] to describe arbitrary systems, but for our analysis of the phosphorylation-based insulator, we will suppose in the native or uninsulated system we have two modules S_1 and S_2 where the dynamics S_1 are given as:

$$\frac{dx_1}{dt} = f_1(x_1, u_1, r), \quad y_1 = Y_1(x_1, u_1, r),$$

and the dynamics of S_2 are given as



Supplementary Figure S4 A schematic illustrating the structure and function of the phosphorylation-based insulator. In a natural uninsulated setting, upon interconnection of an upstream system (S_1) and downstream system (S_2), a "retroactivity signal" comes into existence that may alter the dynamics of the upstream system. However, by inserting an insulating device between the two, retroactivity to the upstream system is abolished and the insulating device is tuned to minimize the impact of retroactivity on its own internal dynamics.

$$\frac{dx_2}{dt} = f_2(x_2, u_2) = f_2(x_2, y_1),$$

$$y_2 = Y_2(x_2, u_2) = Y_2(x_2, y_1),$$

$$r = R(x_2, u_2).$$

Notice that due to the nature of interconnections in biology, the states x_1 of S_1 may have overlap with the inputs u_2 of S_2. In particular, even though y_1 may also be a state in S_1, $y_1 = u_2$ is the input for S_2. Additionally, we will consider systems S_1 and S_2 to be interconnected only to each other. Thus, those familiar with the work of [68] will note that we have omitted the retroactivity to the input signal for S_1 and the retroactivity to the output signal of S_2. The retroactivity signal we wish to center our attention on is r, which describes the retroactivity to the output signal for S_1, or viewed from the perspective of S_2, the retroactivity to the input signal for S_2.

The purpose of the phosphorylation-based insulator is to insulate systems S_1 and S_2 from the retroactivity effects that arise from their interconnection. Mathematically, the insulator can be considered as a separate system S_I with the express purpose of eliminating retroactivity to the output y_1 (and equivalently retroactivity to the input u_2). In the literature, the insulator is inserted as a separate system in between S_1 and S_2 — we will adopt the same paradigm. Thus, the dynamics of the insulated system are given as

$$\frac{dx_1}{dt} = f_1(x_1, u_1), \quad y_1 = Y_1(x_1, u_1),$$

and the dynamics of S_I (denoted S_{Ins} in Figure S3) are given as

$$\frac{dx_I}{dt} = f_I(x_I, u_I, r),$$

$$y_I = Y_I(x_I, u_I, r),$$

$$\frac{dx_2}{dt} = f_2(x_2, u_2) = f_2(x_2, y_1),$$

$$y_2 = Y_2(x_2, u_2) = Y_2(x_2, y_1),$$

$$r = R(x_2, u_2).$$

Notice the former retroactivity of the input u_2, referred to as r, acts as a retroactivity to the output of the insulator module S_I . At the same time, notice that the system dynamics of S_I are such that the upstream system S_1 is insulated from the effects of r; that is, no retroactivity signal maps from the insulator to S_1. In this way, the dynamics of S_I are structured to insulate the upstream module S_1 from the effects of the downstream module S_2.

The key to attenuating the retroactivity r is a design strategy of high gain coupled with negative feedback. This principle is borrowed from the design of electronic amplifiers, where retroactivity is made negligible by a theoretically infinite amplification gain and equally large negative feedback gain. This principle can be motivated using a simple linear systems model. Consider the output function Y_I of the insulator S_I, and suppose that r enters as an additive disturbance in Y_I. We suppose that negative feedback is implemented on the output y_I with gain K and that G is a transfer function describing the insulator S_I. Then supposing we can write the

closed loop (insulator) dynamics of y_I in the Laplace domain, we have the following expression from [68]:

$$\widehat{y_I}(s) = G(s)[\widehat{u_I}(s) - K(s)\widehat{y_I}(s)] + \widehat{r_I}(s),$$

which can be written as

$$\widehat{y_I}(s) = \frac{G(s)}{1 + K(s)G(s)}\widehat{u_I}(s) + \frac{1}{1 + K(s)G(s)}\widehat{r_I}(s).$$

By increasing the gain of either G or negative feedback gain K, we can render the contribution from the retroactivity r negligible. Finally, as we increase the system gain of S_I, namely the gain of G, the signal $\widehat{y_I}(s)$ tends towards $\widehat{u_I}(s)/K$. So far, these observations only impose high-level design specifications on the phosphorylation based insulator. In [68], these design specifications are shown to be satisfied by two types of insulators: a transcriptional feedback insulator and a phosphorylation-based insulator. Our paper focuses on modeling and characterization of the latter in the TXTL system.

Chapter 3

EXPRESSING BIOLOGICALLY ACTIVE MEMBRANE PROTEINS IN A CELL-FREE TRANSCRIPTION-TRANSLATION PLATFORM

Abstract:

Cell-free transcription-translation platforms have been shown to be essential in basic synthetic biological circuit prototyping. From the synthetic biology's point of view, the ability to use membrane proteins in biocircuits in cell-free transcription-translation systems is critical, considering the fact that histidine kinases, G-protein coupled receptors (GPCRs) and other important biosensors are all membrane proteins. Previous studies have expressed membrane proteins in cell-free systems with the help of detergents, liposomes, or nanodiscs, but have not demonstrated the ability to prototype circuit behavior for the purpose of testing more complex circuit functions involving membrane-bound proteins. Built on previous efforts, in this work we demonstrated that we could co-translationally express soluble and active membrane proteins in our cell-free TX-TL platform with membrane-like materials. We first tested the expression of several constructs with β1 and β2 adrenergic receptors in TX-TL and observed significant insoluble membrane protein production. And with the addition of lipoprotein-based membrane-like material – nanodiscs, we were able to make soluble and active membrane proteins which exhibited positive binding to their ligands carazolol and norepinephrine via both fluorescence assay and Surface Plasmon Resonance (SPR) binding assay. Our results suggest that it is

feasible to use cell-free systems to prototype synthetic biocircuits involving single chain membrane proteins without extra procedures, leading us one step closer to testing complex membrane protein circuits in cell-free environment.

3.1 Introduction

Cell-free transcription-translation systems have shown to be extremely useful in synthetic biological circuit prototyping [23, 24]. Typical cell-free transcription-translation systems are based on S30 *E. coli* extract [38] and there have been many different versions. The specific version used in this paper referred to as "TX-TL", has been optimized for prototyping synthetic biocircuits [20, 21]. Unlike other cell-free protein expression systems, including the PURE system, which are based on bacteriophage transcription by supplementing bacteriophage RNA polymerases to the crude cytoplasmic extracts [69], TX-TL has been shown to be suitable for complex biochemical systems [20, 22-24].

Membrane proteins are of great importance for proper functioning cells and organisms. They lay the foundation for many biosensors and signal pathways in cells [70]. Some membrane proteins act as ion channels to transport ions across membranes [71]; others make up the essential parts of sensory system which are responsible for cell communication [72], while membrane enzymes catalyze important chemical reactions near membranes [73]. In addition, membrane proteins are the most important drug targets, for example more than half of therapeutics for treatment of various modalities ranging from cancer to cardiovascular diseases target membrane proteins [74].

Membrane proteins are difficult targets as compared to soluble proteins because of the challenges associated with their expression, solubilization, and stabilization.

Typical studies of membrane proteins rely on proteins produced from cells and solubilized cell membrane using detergents, liposomes, or other membrane-like materials. Although this approach may have advantages in terms of protein yield, it is not well-suited for high-throughput assays since for each construct transformation, cell growth, lysis, membrane solubilization, and purification are involved [75]. Furthermore, these techniques are not suitable for biocircuits prototyping either because the membrane proteins have to be in functional conformations right after expression for the circuits to work without further solubilization or purification processes. As a result, the best way to approach it would be to express membrane proteins *in vitro* in presence of a membrane-like material. There have been several studies on using detergents, liposomes, or nanodiscs to help solubilize and stabilize membrane proteins generated in cell free system [75-81], but there are limitations of these methods in translating results to circuit prototyping in TX-TL.

Here, we demonstrated that membrane proteins can be directly integrated into future synthetic biocircuits for performing more complex functions (an additional study on a potassium channel membrane protein can be found in Supplementary Materials *"Expressing a chimera potassium channel membrane protein KcsA-Kv1.3 in TX-TL"*). Among all the important membrane proteins, we picked G protein coupled receptors (GPCRs), β2 (and β1) adrenergic receptors (β2AR/ β1AR) as model proteins [82, 83]. With more than 900 members, GPCRs are one of the most important and the largest integral membrane proteins family in human cells and the most important clinical drug targets as they play important roles in many

physiological functions and implicated in many diseases [84, 85]. GPCRs all share the same topology – seven transmembrane α-helices, and they are thought to function in monomeric form [86], although there have been studies indicating their dimerization [87, 88]. We used TX-TL platform in combination with nanodisc for expressing β2AR/β1AR and subsequent *in situ* stabilization (Figure 1A). Furthermore, biological activity of these proteins was confirmed by testing their bindings to the ligands.

3.2 Results and Discussion

We ordered gene synthesis services and then used golden gate assembly [89] to make three different GPCR constructs with three different variants, which share the same backbone, promoter, ribosome binding site, and terminator (Figure 1B); Table 1 lists the difference in coding sequences corresponding to various adrenergic receptor constructs. All three constructs also shared the same superfolder GFP (sfGFP) fusion protein topology, which was tagged with 6xHis tag at the C-terminal of the protein. sfGFP is used for monitoring and quick estimation of target protein expression level

GPCR Protein	Construct	Protein Size
β1-AR-ts: 3ZPR Thermostabilized turkey β1 adrenergic receptor	pSG73: protein-sfGFP-His^6	65kD (35kD minus sfGFP)
β2-AR wild type: ADRB2_HUMAN β2 adrenergic receptor	pSG74: protein-sfGFP-His^6	75kD (45kD minus sfGFP)
β2-AR-T4L: 2RH1 β2-adrenergic receptor/T4-lysozyme chimera	pSG75: protein-sfGFP-His^6	85kD (55kD minus sfGFP)

Table 1 Information of GPCR constructs used in experiments. pSG73 and pSG75 have the corresponding PDB number noted. Protein size of each fusion protein is also listed.

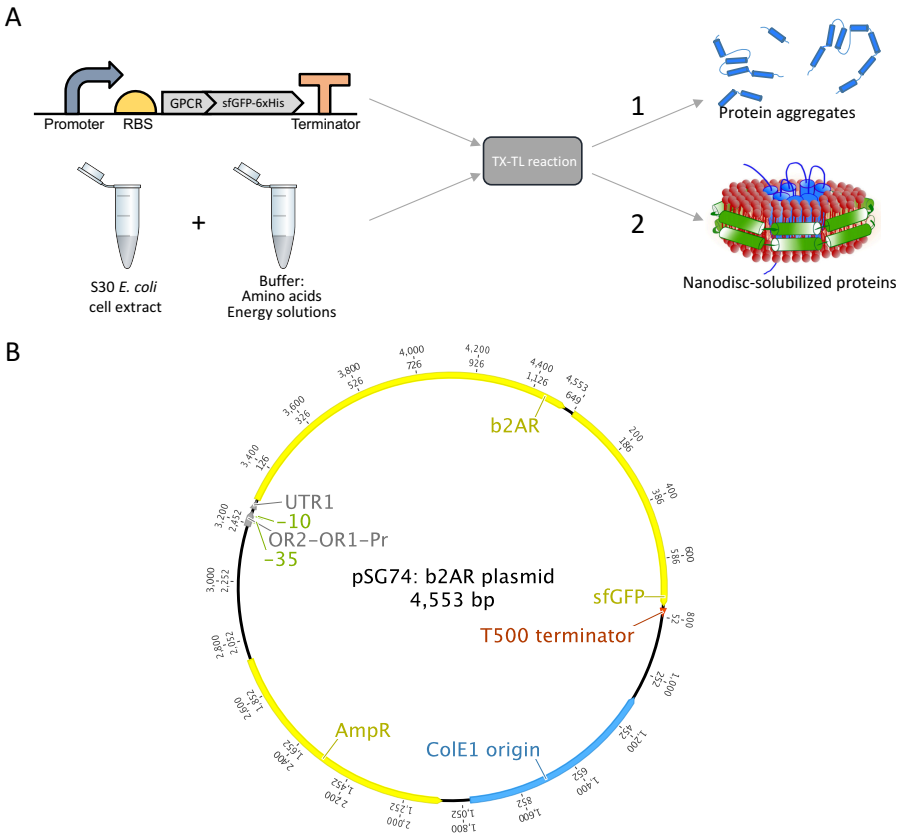


Figure 1 Illustration of the TX-TL experimental setup and a plasmid map of the constructs. **A**: A simplified illustration of the TX-TL experimental setup with DNAs, cell extracts and buffers. TX-TL reactions designed for membrane protein expression can yield two different outcomes: 1) protein aggregates without any membrane-like materials; 2) nanodisc-solubilized proteins when nanodiscs are added. **B:** Circular plasmid map of pSG74. pSG74 and pSG75 share the same features as pSG73.

during and at the end of TX-TL. Whereas, 6xHis tag was used to detect proteins in Western, to capture the protein for binding assays, and for affinity purification, if necessary.

One of the advantages of TX-TL expression platform is that we can use either linear DNA or plasmid DNA for expression in TX-TL [22]. Because of that, we can

implement fast construct prototyping in TX-TL by ligating parts together and amplifying the linear DNAs with PCR, avoiding cloning the linear fragments into plasmid. To express these constructs in TX-TL, we simply add linear DNAs of these constructs to TX-TL reaction mixes. The iteration of each prototyping test using linear DNAs in TX-TL can take as less as 12 hours, much faster compared to traditional approaches, which can take around days to weeks for one iteration.

We first estimated the expression level of pSG73-75 in TX-TL by measuring the fluorescence of sfGFP which was fused at the C-terminal of β2AR or β1AR, using a plate reader at 485 nm (absorbance)/525 nm (emission). All the linear DNA constructs showed a GFP fluorescence signal, indicating successful expression of the fusion proteins in TX-TL (Figure 2A). Different constructs, despite having exactly the same promoter, ribosome binding site, fusion protein framework, and DNA concentration, showed different expression levels of sfGFP, especially pSG74. This suggested that the difference in coding sequences could affect transcription and translation level (detailed descriptions of "*The trials and errors of the expressions of β1AR/β2AR proteins in TX-TL*" can be found in Supplementary Materials. Another observation was that increasing linear DNA concentration could help increase the fusion protein expression, but this increase was limited by TX-TL resources and/or toxics accumulated in batch mode, as shown elsewhere [37]. Setting up TX-TL reactions in dialysis systems is likely to improve the protein expression level.

After testing the fusion protein expression level, we explored the possibility of stabilizing hydrophobic membrane protein by providing a membrane mimic into the

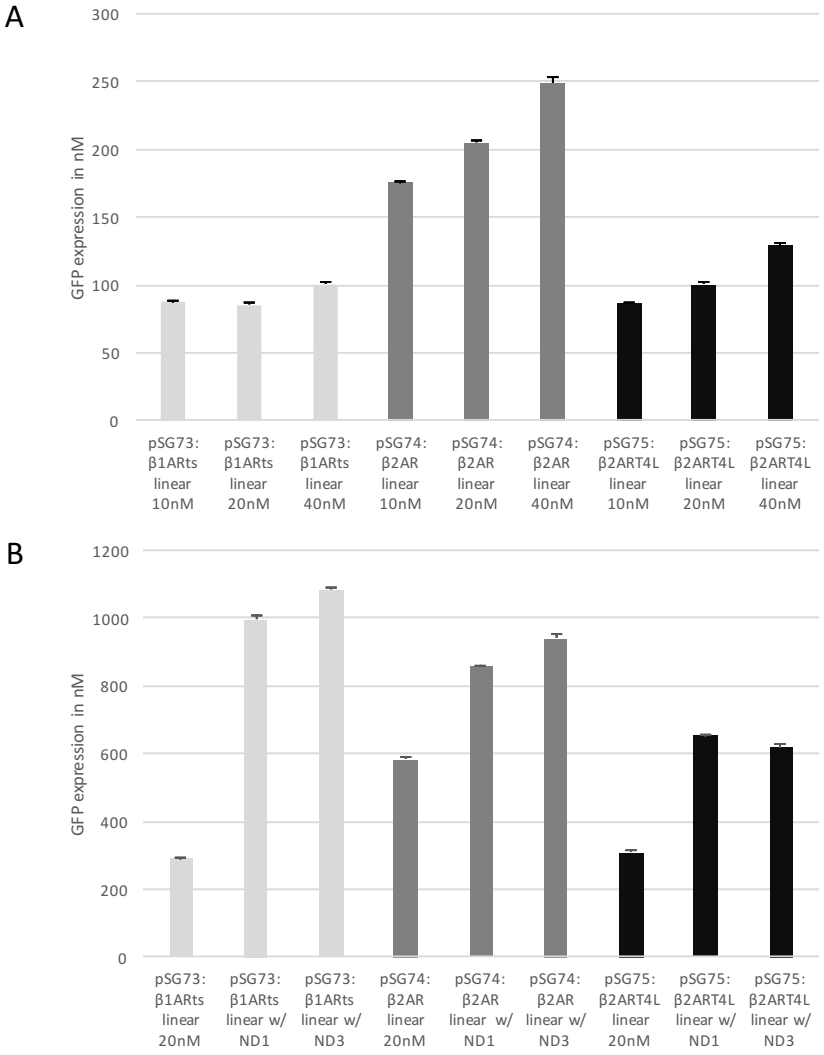


Figure 2 Expression of pSG73-75 constructs in TX-TL measured by GFP fluorescence. **A:** End point measurement of GFP expression from 10nM, 20nM and 40nM linear DNA of pSG73, pSG74 and pSG75 in 10μL TX-TL. **B:** End point measurement of GFP expression from 20nM linear DNA of pSG73-75 with or without 24μM nanodiscs in TX-TL. ND1 is MSP1D1-DMPC and ND3 is MSP1E3D1-DMPG. Measurements were done at 29°C in BIOTEK Synergy H1 Hybrid Multi-Mode Microplate Reader using ex485nm/em525nm.

reaction. Detergent micelles, liposomes, and nanodiscs are commonly used to provide a hydrophobic environment to membrane proteins [90]. We observed that the majority of detergents were detrimental to the TX-TL reaction comprising

membrane proteins (Supplementary Figure S1). Conversely, reconstituting protein into the liposome resulted in a poor yield of the folded protein, which could be attributed to the liposome's closed topography. Therefore, we employed a nanodisc, a lipid-protein complex composed of lipids constrained by a membrane scaffold protein (MSP). The nanodisc provides a robust platform with two-dimensional topography for reconstituting TX-TL expressed membrane protein. We chose two of the most commonly used nanodiscs: MSP1D1-DMPC (ND1) ~10nm diameter, and a slightly larger version MSP1E3D1-DMPG (ND3) ~13nm in diameter [75, 91].

We repeated the TX-TL expression experiment in the presence/absence of ND1 or ND3. As shown in Figure 2B, the presence of the nanodisc improves the TX-TL efficiency, as indicated by enhanced GFP fluorescence. The nanodisc does not have any intrinsic fluorescence, and therefore the increased fluorescence should be arising from the fusion protein. GPCR-sfGFP fusion protein precipitated in the TXTL reaction mix without nanodisc. However, in the presence of nanodisc, the supernatant of the reaction mixture exhibited increased GFP fluorescence signal, indicating that nanodiscs likely help stabilize fusion membrane proteins in TX-TL, and keep the folded fusion protein in supernatant.

To further confirm that the target fusion proteins were expressed in TX-TL and nanodiscs could help solubilize GPCR-sfGFP fusion proteins, we ran SDS-PAGE of TX-TL samples and western blots with an anti-His6 antibody. First, we ran the whole reaction samples of three different constructs at three different concentrations (Figure 3A). As illustrated on the blot, samples that had one of the three constructs

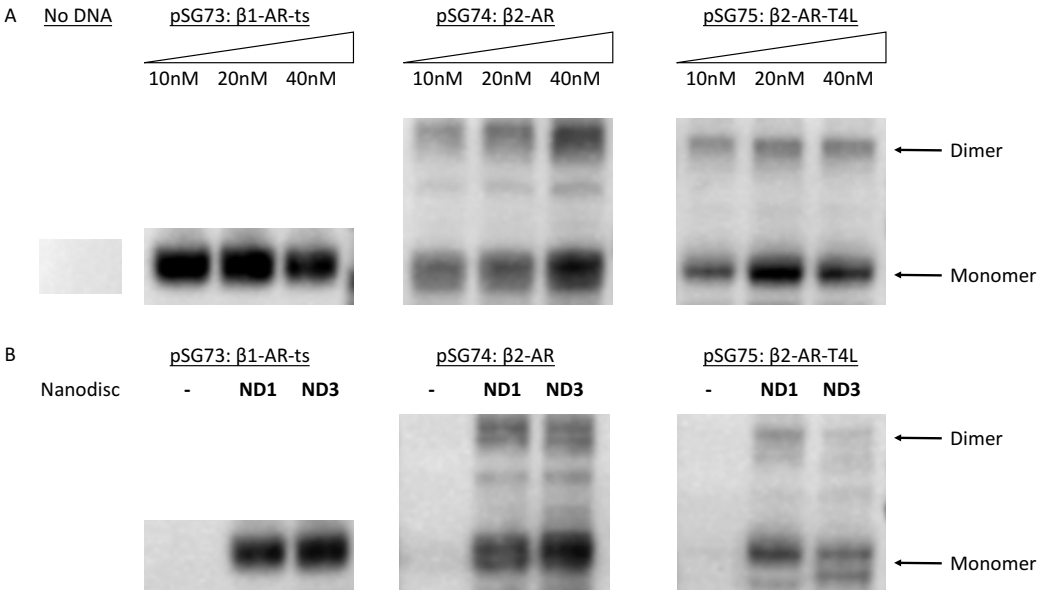


Figure 3 Western blots of TX-TL reactions. **A:** Results of whole reaction TX-TL samples. After measuring the fluorescence of the TX-TL reaction, they were used for western blot using the method in Materials and Methods. TX-TL reaction with no DNA was used as negative control and three different concentrations of three different constructs were run under the same condition. **B:** Results of only supernatant from TX-TL samples. Only soluble samples were run on this blot. Supernatants without any nanodiscs were run at the same condition as ones with either ND1 or ND3. Full gels can be found in Supplementary Figures S7 and S9 in "*The trials and errors of the expressions of β1AR/ β2AR proteins in TX-TL".*

showed significant detection of His-tagged proteins, and protein size was confirmed on the blot using a ladder. In contrast, the negative control sample without DNA had no His-tagged signal. Additionally, there were not significant differences between different concentrations in pSG73 and pSG75, suggesting mass-transfer nutrients limitation and toxic accumulation in the TX-TL reaction as discussed earlier. Another interesting observation was that pSG73-β1AR-ts did not show dimerization but pSG74 and pSG75, which were both β2AR based fusion proteins, showed strong dimer bands on the blot. The presence of GPCR dimerization is consistent with previous reports [88, 92].

We further tested the presence of ND by spinning down the TX-TL reaction mix and took supernatant only to run the western blot. All the insoluble proteins would precipitate out and ended up in the pellet after centrifugation. Only soluble proteins would be in the supernatant and can be detected in the western blot. In Figure 3B, we had three different constructs and each of them had three different experimental conditions: no nanodisc, with ND1 (MSP1D1-DMPC), and with ND3 (MSP1E3D1-DMPG). All target proteins generated in TX-TL precipitated without nanodiscs and left in the pellet (confirmed by running pellet on western blot, data not shown). On the contrary, when either ND1 or ND3 was added into the reaction, we saw significant bands of target membrane proteins by the Western, suggesting that these hydrophobic membrane proteins became soluble with help from nanodiscs.

So far, we have demonstrated that target membrane proteins produced in TX-TL reactions can be solubilized and stabilized with the *in situ* presence of nanodiscs. However, protein association with the nanodisc does not insure that these proteins are biologically active. To further test whether these soluble proteins were active, we designed two binding assays: fluorescence-based carazolol binding assay and surface plasmon resonance (SPR) based norepinephrine binding assay.

Although there is no need to purify the protein for circuit prototyping, purification is required to confirm the binding activity of TX-TL expressed protein to avoid interference caused by *E. coli* endogenous proteins in the TX-TL reaction mix. We use Ni-NTA affinity chromatography to purify His-tagged target protein.

Fluorescence-based carazolol binding assay uses (S)-carazolol, a derivative of the potent β blocker carazolol with fluorescence properties (ex633nm/em650nm) [93]. This ligand can be used as a fluorescence tracker for β2AR binding activity. The detailed experimental setup is described in Materials and Methods. Briefly, purified membrane proteins were incubated with or without carazolol for one hour. These samples were further dialyzed against 100X volume buffer overnight. Subsequently, all the samples were concentrated down to their starting volume and fluorescence signal was measured in a plate reader. Green fluorescence (GFP) indicates the amount of fusion proteins in the samples and red fluorescence represents the amount of carazolol bound to target membrane proteins as an indication of protein activity. We used two different ND1 (MSP1D1-DMPC) in this assay: 1) HisND uses 6xHis tagged MSP1D1 and was purchased from Cube Biotech; 2) BiotinND uses biotinylated MSP1D1 and was made in house.

Results from fluorescence binding assays showed that there was not much difference in the green fluorescence signal between the samples (Supplementary Figure S2). To simplify the data, we normalized the fold change of carazolol red fluorescence using the fold change of GFP fluorescence to calculate the relative carazolol bound to each GPCR-sfGFP protein (Figure 4A). As we can see in Figure 4B, we observed significant associations of carazolol to every fluorescence unit of GPCR-sfGFP protein. We attribute the higher carazolol signal on HisND reconstituted GPCR to the presence of empty ND after Ni-NTA affinity chromatography purification.

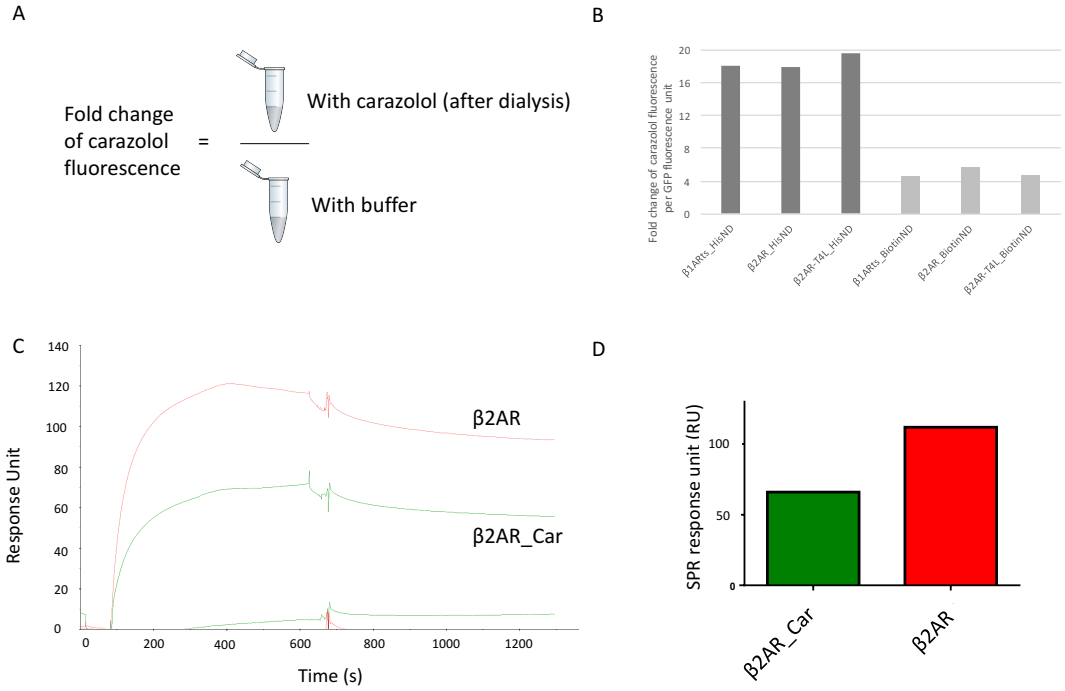


Figure 4 Binding assays of TX-TL made β1 and β2 adrenergic receptor. **A:** Illustration of how we calculated the fold change of carazolol red fluorescence. The number was the ratio of red fluorescence signal from β1 and β2 adrenergic receptor samples incubated with carazolol to the ones without carazolol. A detailed experimental method is in Materials and Methods. All samples were dialyzed against buffer without carazolol overnight to remove non-binding carazolol. The exact same method was also used to calculate fold change of GFP fluorescence. **B:** Bar chart of the fold change of carazolol red fluorescence normalized with GFP fluorescence from the same samples. The number was calculated by dividing the fold change of carazolol red fluorescence by the fold change of GFP fluorescence. What we have here essentially indicates how much carazolol is associated with each GPCR-sfGFP protein. **C:** Response curves of β2 adrenergic receptor proteins binding to SPR surface coated with norepinephrine. Red curve is the response signal from β2-AR sample without carazolol. Green curve is the response signal from sample incubated with 1μM carazolol. The bottom curves are the corresponding control channels. **D:** Bar chart of the end point of the response curves in C.

One limitation of fluorescence-based assay is that non-specific binding of carazolol to lipids surrounding membrane protein could cause higher signal-to-noise ratio. To overcome this, we developed a SPR-based binding assay using norepinephrine, which is a partial agonist for β2AR. Since norepinephrine shares the same binding pocket with carazolol [94], we tethered norepinephrine to the surface

via amide bond between its primary amine and the carboxyl group on the surface. We hypothesized that carazolol could be used as a competitor for β2AR to norepinephrine binding on SPR. As shown in Figure 4C, SPR binding response of β2AR decreased to ~60% when it was incubated with 1μM carazolol, indicating specific interaction between β2AR and its binding partners.

To summarize, we have demonstrated that 1) membrane protein can be expressed in TX-TL at analytical scale; 2) presence of nanodisc during TX-TL reaction facilitates folding and solubilization of single chain membrane protein; 3) fluorescence and SPR binding assays were developed to demonstrate specific interaction between small-molecule and nanodisc-stabilized membrane protein.

3.3 Conclusions

In this work, we tested proteins from the GPCR family in our cell-free transcription-translation (TX-TL) system, which would be ideal for synthetic biocircuit prototyping. We expressed β1AR/ β2AR in TX-TL with nanodiscs and were able to show that not only these membrane proteins are soluble in TX-TL with nanodiscs, but they were also active. Nanodiscs are co-translationally associated with membrane proteins without extra processing, which enables one pot prototyping in TX-TL for testing a complete biocircuit with membrane protein of interest involved.

We intend to optimize our binding assays and perform competitive binding experiments to test the stringency of this assay. We envision that GPCR co-expressed

with G protein can be expressed by TX-TL to test the biological circuit including signal transduction.

Additionally, histidine kinases, which phosphorylate corresponding response regulators, can activate downstream transcription and translation. We have started tested some hybrid histidine kinases and have seen promising results. Our goal is to prototype a logic biocircuit with membrane enzyme in it, expanding TX-TL platform to broader topics.

Materials and Methods:

Plasmids and linear DNAs:

DNA and oligonucleotides primers were ordered from Integrated DNA Technologies (IDT, Coralville, Iowa). Plasmids in this study were designed in Geneious 8 (Biomatters, Ltd.) and were made using standard golden gate assembly (GGA) protocols. BsaI-HF (R3535S) enzyme used in GGA was purchase from New England Biolabs (NEB). Linear DNAs were made by PCRing protein expression related sequences out of GGA constructs using Phusion Hot Start Flex 2X Master Mix (M0536L) from NEB.

TX-TL reactions:

TX-TL reaction mix was set up according to previous JOVE paper [21]. Briefly, TX-TL extract and buffer were mixed together with calculated linear DNAs or plasmids with or without nanodiscs. Reaction volumes varied from 10μL (initial screening) to 1mL (protein purification and analysis).

Gel and western blot:

Gels used in this work were Bolt 4–12% Bis-Tris Plus Gels from ThermoFisher Scientific. Running buffer was Bolt MES SDS Running Buffer. Gels were run without reducing agents. Protein samples were mixed with LDS sample buffer before loading into gels. iBlot 2 Gel Transfer Device and iBlot Nitrocellulose Regular Stacks were used for transfer proteins from gel to membrane. Membrane was then

transferred to iBind device and incubated with Penta-His HRP Conjugate in 1:500 dilutions. Blots were detected using SuperSignal Chemiluminescent HRP Substrates from ThermoFisher Scientific.

Protein purification:

TX-TL reaction mix was first spun @14,000g for 10min at 4°C. Supernatant was then transferred to buffer-equilibrated HisPur Ni-NTA Spin Purification column (ThermoFisher Scientific) and incubated with shaking for 1h, and then spun down the flow through @2000g for 2min and washed with nanodisc buffer (20 mM Tris pH 7.4, 0.1 M NaCl) added with 20mM imidazole three times. Elution was done by adding elution buffer (20 mM Tris pH 7.4, 0.1 M NaCl, 250mM imidazole) for three times 2 column volume. Proteins were then concentrated using Amicon Ultra Centrifugal Filter Units (Millipore) with Ultracel-30 membrane.

Fluorescence-based carazolol binding assay:

Carazolol was purchased from Abcam (S)-Carazolol Fluorescent ligand (Red) ab118171. Each purified protein was first divided into two equal volume samples. To one was added 100nM carazolol (dissolved in water) and to the other was added the same volume of water. They were incubated at 4°C for 1h before being transferred to mini 10k MW D-Tube Dialyzers (Millipore) and dialyzed against 100x volume of nanodisc buffer for overnight. Then dialyzed samples were concentrated using Amicon Ultra Centrifugal Filter Units (Millipore) with Ultracel-30 membrane to the starting volume. Samples were then put into a Biotek synergy H1 hybrid multi-mode

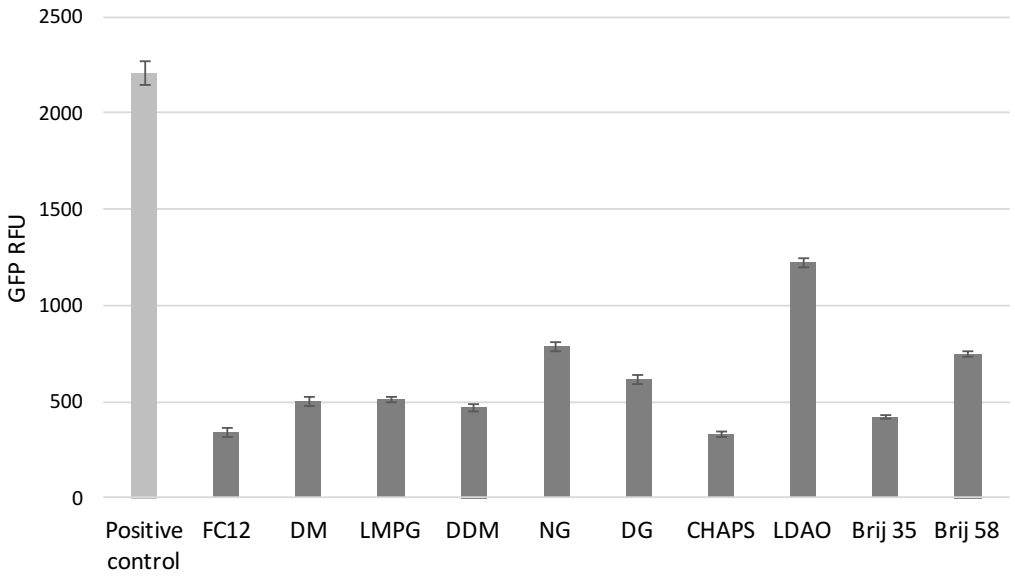
microplate reader and measured for GFP fluorescence (ex485nm/em525nm) with gain 61 or gain 100 or Red fluorescence (ex633nm/650nm) with optimal gain. GFP fluorescence was converted to nM using calibration data from purified GFP protein.

Surface Plasmon Resonance (SPR) based norepinephrine binding assay:

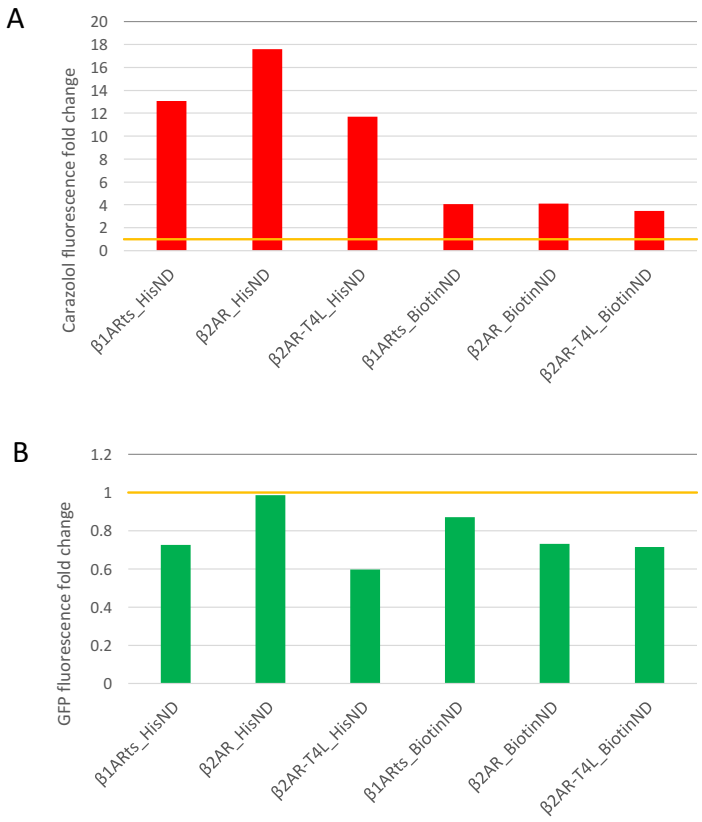
GE Biacore T-200 SPR system was used for SPR experiment. Gold plated chip was first immobilized with norepinephrine and then washed away extra chemical. The equilibrating buffer and running buffer were both the nanodisc buffer. There were four channels on one chip. Two were used as experimental channels and the remaining two were used as negative controls to provide background response from buffer. One sample was incubated with 1μM carazolol for 1h at to test binding specificity and the other sample was incubated with same volume of water. Samples were then loaded and flowed through corresponding experimental channels and response curves were recorded.

Supplementary Materials

Supplementary Figures



Supplementary Figure S1 Bar chart showing the impact of various detergents on the expression of GFP in TX-TL. All samples had 10 nM of pCon-GFP DNA in it. Except for the positive control sample, reactions were added with the detergents marked under the bars. For detergents FC12 (dodecylphosphocholine), DM (n-decyl-β-D-maltoside), NG (n-nonyl-β-D-glucoside), DG (n-decyl-β-D-glucoside) and CHAPS, 2X of the critical micelle concentration (CMC) was used. For detergents LMPG, DDM (n-dodecyl-β-D-maltoside), and LDAO (lauryldimethylamine N-oxide), 5X of CMC was used. For Brij 35 and Brij 58, 10X CMC was used.



Supplementary Figure S2 A: Bar chart of carazolol fluorescence fold change. The fold change is the ratio of red fluorescence signal from samples incubated with carazolol to the ones without carazolol. A detailed experimental method is described in Materials and Methods. All samples were dialyzed against buffer without carazolol to remove non- binding carazolol. **B:** Bar chart of GFP fluorescence fold change with the same samples from A. The fold change is the ratio of green fluorescence signal from samples incubated with carazolol to the ones without carazolol.

SM1 The trials and errors of the expressions of β1AR/ β2AR proteins in TX-TL

We made 4 different beta adrenergic receptor constructs to test the expression in TX-TL:

1. pSG72: Beta-1 adrenergic receptor (β1AR): P08588|ADRB1_HUMAN Beta-1 adrenergic receptor. Codon optimized for expression in TX-TL.

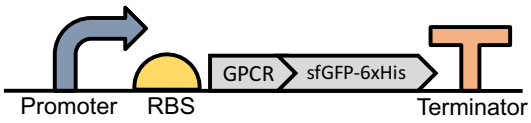
2. pSG73: β1ARts: 3ZPR. Thermostabilized turkey beta-1 adrenergic receptor. Codon optimized.

3. pSG74: β2AR: P07550|ADRB2_HUMAN beta-2 adrenergic receptor. Codon optimized.

4. pSG75: β2ART4L: 2RH1. Beta-2-adrenergic receptor/T4-lysozyme chimera. Codon optimized.

All four genes were inserted into a backbone with OR2OR1-Pr as the promoter, UTR1 as the ribosome binding site, fused to a sfGFP gene, ColE1 as the replication origin, and Amp as the antibiotic selection marker (Figure 1B).

We also made linear DNAs of each construct as shown in Supplementary Figure 3 (except pSG72 where PCR failed). The pSG73 linear DNA map is used below for illustration purposes. The reason to use linear DNAs is that it is easier to make high concentration materials in a short time (3 hours using just PCR). The expression is almost as good as plasmids with protection protein gamS at present.

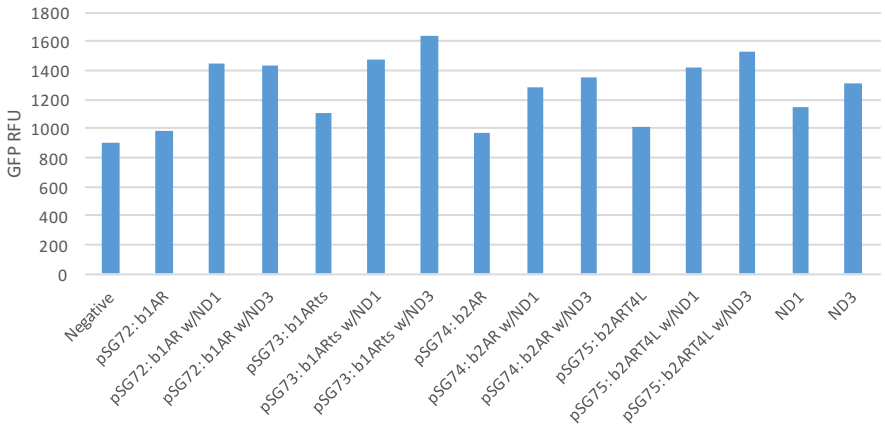


Supplementary Figure S3 Illustration diagram of the linear DNAs used for expression in TX-TL. They all share the same backbone with the same promoter, ribosome binding site, and terminator. They are all fused to a sfGFP-His tag.

1 First expression test using plasmids

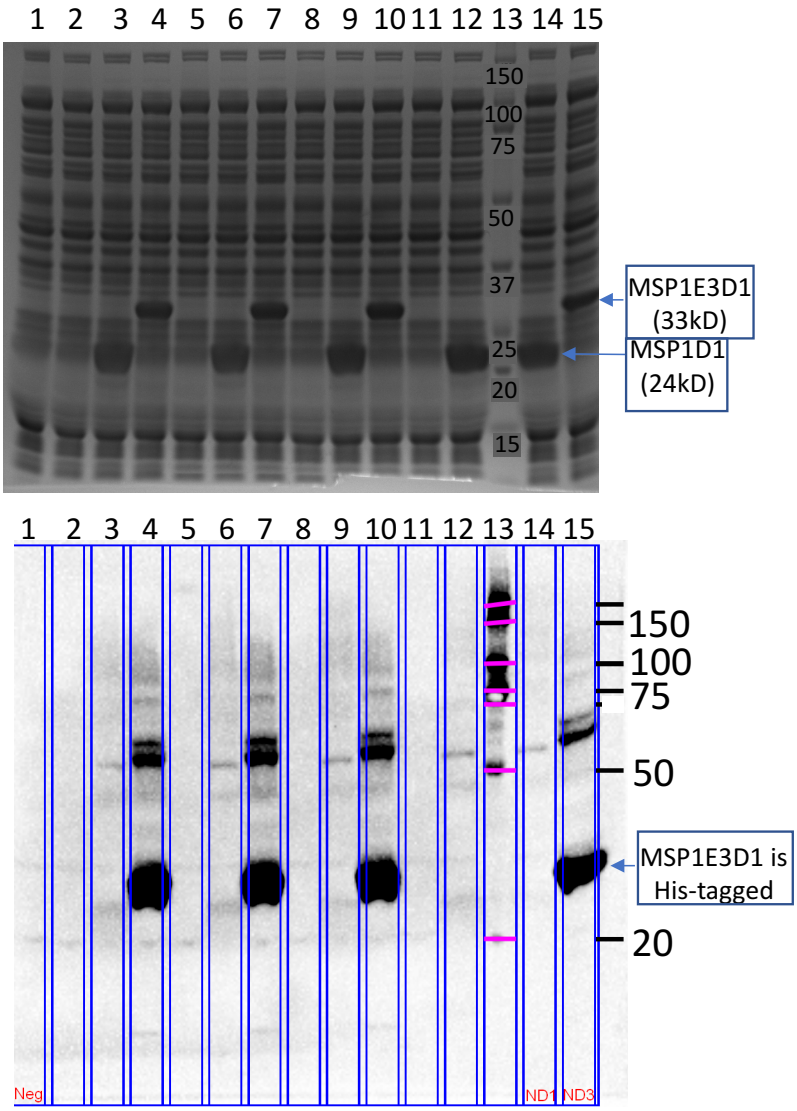
For the first expression test, we put four plasmids pSG72-75 into TX-TL reactions with concentrations ranging from 4 nM to 7 nM with or without 40 nM MSP1D1-His_DMPC (ND1) and MSP1E3D1-His_DMPG (ND3) and incubated the reaction mixes at 29°C for overnight.

Because we have sfGFP fused to the target proteins, we used a plate reader to record real-time GFP fluorescence. However, the results were not promising.



Supplementary Figure S4 Bar chart of the expression from constructs pSG72, 73, 74, and 75 in plasmids with or without nanodiscs, ND1 or ND3. The negative control was a TX-TL reaction without any DNA and showed the background fluorescence of the reaction mix. The Y axis was GFP relative fluorescence unit measured by a Biotek plate reader.

As we can see in Supplementary Figure S4, there was no significant protein expression above the negative control. We suspected that the DNA concentrations were too low to generate enough GFP signal. We then took the whole reaction mix (1 μL) and ran a SDS-PAGE gel (then stained with Coomassie Blue); we also used the same whole reaction mix (3 μL) and ran a western blot (Supplementary Figure S5). As we can see in the top panel, there were significant amounts of MSPs from the two different nanodiscs that we used in the experiment. But we did not see any other protein bands that were not in the negative control. On the western blot, we saw very strong bands from nanodisc MSP1E3D1, but not so much from nanodisc MSP1D1. And, consistent with the plate reader results, we did not see any of the target protein bands on the western blot.

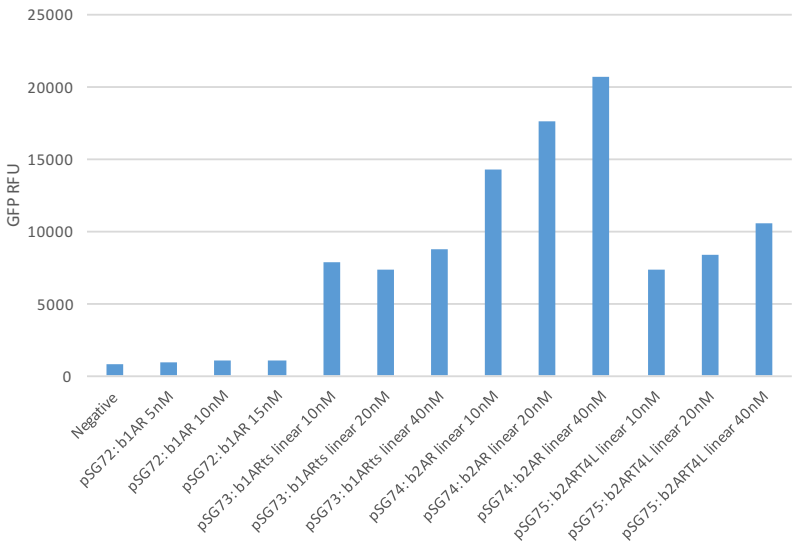


1	2	3	4	5	6	7	8
Negative	pSG72: b1AR	pSG72: b1AR w/ND1	pSG72: b1AR w/ND3	pSG73: b1ARts	pSG73: b1ARts w/ND1	pSG73: b1ARts w/ND3	pSG74: b2AR
9	10	11	12	13	14	15	
pSG74: b2AR w/ND1	pSG74: b2AR w/ND3	pSG75: b2ART4L	pSG75: b2ART4L w/ND1	Ladder	ND1	ND3	

Supplementary Figure S5 Top: SDS-PAGE gel results of the TX-TL reaction samples. Middle: Western blot results of the same gel. Bottom: A table describing the samples in each lane.

2 Second test with linear DNAs and pSG72 plasmid

We thought that in the previous trial we did not have high enough of the plasmid concentrations to get strong fluorescence signal from TX-TL reactions. So, we decided to generate more materials by PCRing off those plasmids to get linear DNAs of them. All PCR reactions worked except for pSG72. So, we ended up using pSG72 plasmid and pSG73-75 linear DNAs in this test. The pSG72 plasmid has a concentration range of 5 nM, 10 nM, and 15 nM and the pSG73-75 linear DNAs have a concentration range of 10 nM, 20 nM, and 40 nM. The fluorescence data of the expression of the aforementioned samples are plotted in Supplementary Figure S6.



Supplementary Figure S6 Bar chart of the expression from constructs pSG72 in plasmids, 73, 74, and 75 in linear DNAs without nanodiscs. Three different DNA concentrations were used for each construct. The negative control was a TX-TL reaction without any DNA and showed the background fluorescence of the reaction mix. The Y axis was GFP relative fluorescence unit measured by a Biotek plate reader.

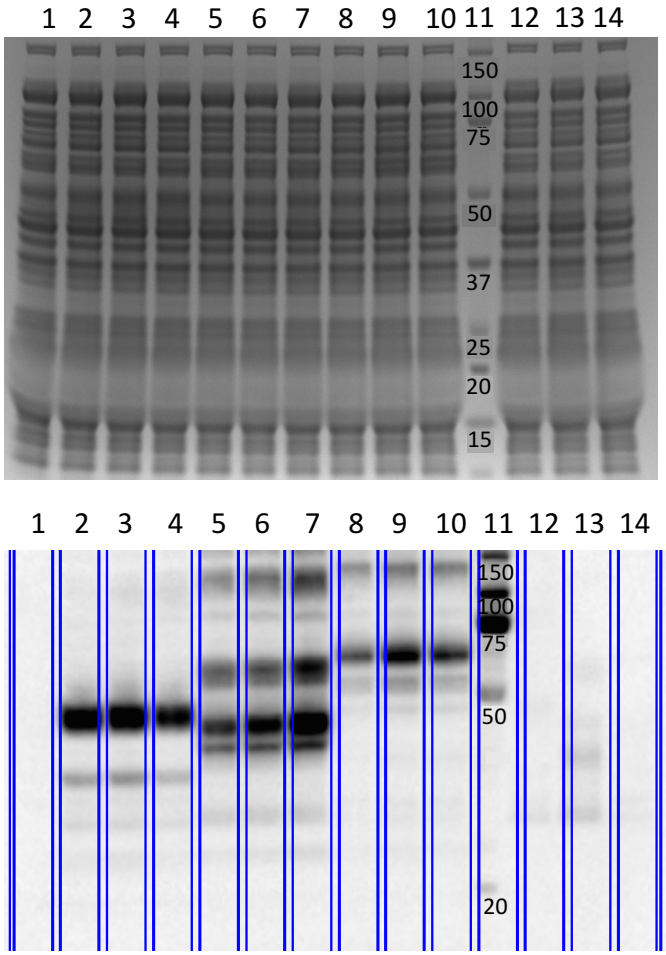
We can draw the following observations from this figure:

1. The plasmid pSG72 still did not work although we increased the concentration to 15 nM. Something was wrong with all the plasmids that we used in this and the previous experiments.

2. Fortunately, all the linear DNAs that we tried worked as expected. Both pSG74 and 75 showed increased fluorescence when we increased the DNA concentrations. Only pSG73 reached maximum expression level at 10 nM.

3. Usually linear DNA has lower expression level compared to the level of the plasmid DNA with the same concentration, meaning 10nM linear DNA should have less fluorescence compared to 10 nM plasmid DNA. We did not observe this in these experiments, which confirmed that there was something wrong with the plasmids. However, those linear DNAs based on these plasmids turned out to be good. The only difference here was that linear DNAs required gamS protection protein in TX-TL; otherwise linear DNAs would be degraded much faster and produce much fewer protein.

The fluorescence data were very encouraging. So, we took 1uL of the whole reaction mixes that showed significant fluorescence signals and ran both SDS-PAGE and western blots. To check whether the beta AR protein was in the precipitate or supernatant, we also spun down the whole reactions at 14,000 rpm for 10 min and took 1 uL of the supernatant and included them in the gels.



1	2	3	4	5	6	7
Negative	pSG73: b1ARts linear 10nM	pSG73: b1ARts linear 20nM	pSG73: b1ARts linear 40nM	pSG74: b2AR linear 10nM	pSG74: b2AR linear 20nM	pSG74: b2AR linear 40nM
8	9	10	11	12	13	14
pSG75: b2ART4L linear 10nM	pSG75: b2ART4L linear 20nM	pSG75: b2ART4L linear 40nM	Ladder	Supernatant: pSG73: b1ARts linear 20nM	Supernatant: pSG74: b2AR linear 20nM	Supernatant: pSG75: b2ART4L linear 20nM

Supplementary Figure S7 Top: SDS-PAGE gel results of the TX-TL reaction samples. Middle: Western blot results of the same gel. Bottom: A table describing the samples in each lane.

In Supplementary Figure S7, the top panel is the SDS-PAGE coomassie blue staining gel. We did not see any significant overexpression band on this gel,

suggesting that although we could detect fluorescence signal from the fusion protein, the amount was not enough to be seen on the less sensitive coomassie blue stain gel. The middle panel is the western blot of the same gel with anti-His antibody. We can make these following observations from the blot:

1. Lanes 2-4 showed significant His tagged protein expression in the whole reaction mix, which was consistent with the fluorescence data. As the DNA concentration increased from 10 nM to 40 nM, we did not see significantly higher protein expression, suggesting it might have reached max level, which was also consistent with the plate reader fluorescence data. There were two significant bands in lanes 2-4. There was a strong band around 50 kD and β1ARts-sfGFP-His was around 65 kD. This had to be β1ARts-sfGFP-His as sfGFP-His was less than 30 kD, which could be the lower weak band.

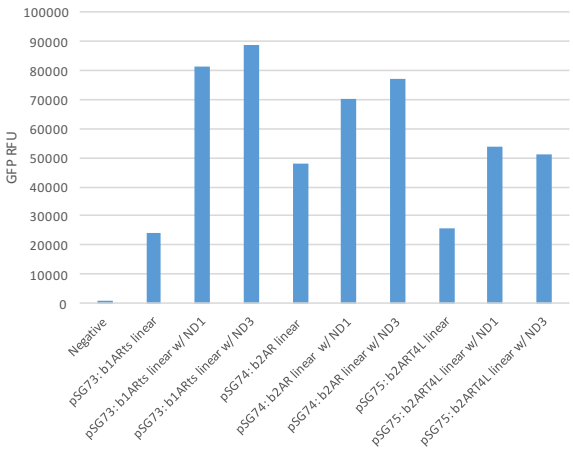
2. Lanes 5-7 showed significant His tagged protein expression in the whole reaction mix as well. As the DNA concentration increased from 10nM to 40nM, we saw significant higher protein expression from lanes 5 to 7, which was also consistent with the fluorescence data. However, these three lanes had more complicated band patterns: two smearing bands over 100 kD, one band between 50 kD and 75 kD and one intense band with a lighter band between 37 kD and 50 kD. β2AR-sfGFP-His should be around 76 kD. Nevertheless, we should have β2AR-sfGFP-His protein in the mix.

3. Lanes 8-10 showed similar results as the two we discussed above. β2ART4L-sfGFP-His should be around 85 kD. We could see one intense band over 100 kD and another one between 50 kD and 75 kD.

4. Lanes 12-14 are critical lanes as they are the supernatant samples of lanes 3, 6, and 9. As we can see very clearly in the blot, lanes 12-14 had almost no significant bands, suggesting the beta AR transmembrane proteins were mostly in precipitate, as we expected.

3 Third test with linear DNA and nanodiscs

Because of the encouraging results in the previous experiment, we decided to try nanodiscs again with 20 nM of the three linear DNAs pSG73-75. The results from the experiment are shown in Supplementary Figure S8.



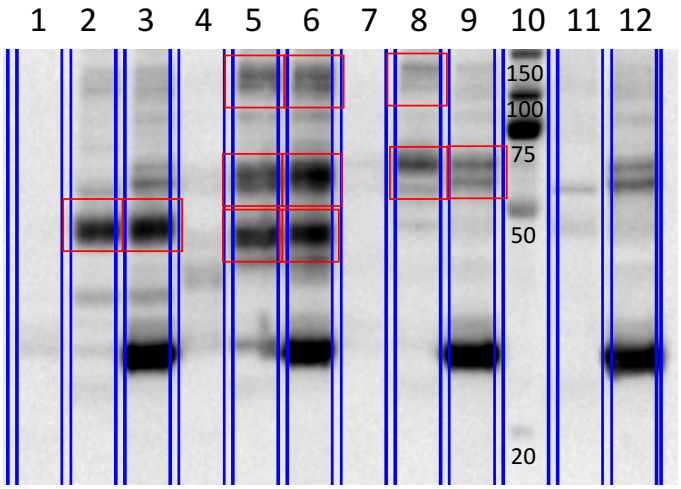
Supplementary Figure S8 Bar chart of the expression from constructs pSG73, 74 and 75 in linear DNAs with or without 40 nM nanodiscs ND1 or ND3. The linear DNA concentration used for all three was 20 nM.

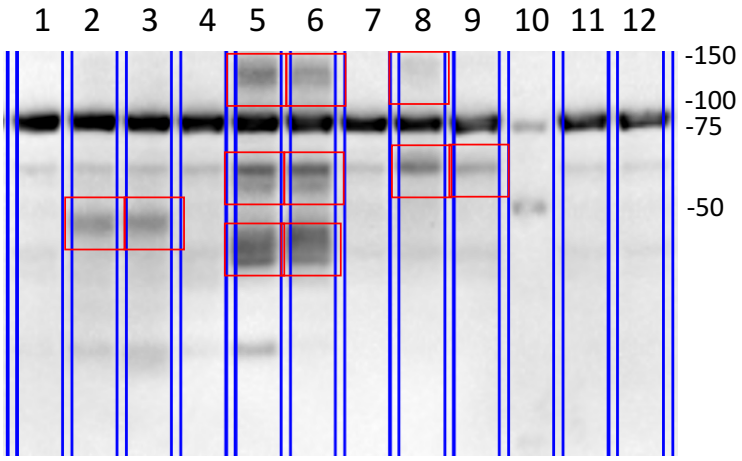
As we expected, all the linear DNAs showed strong expression just like the previous experiment. This time, we had nanodiscs as well in the reaction mix, and having nanodiscs in the reaction significantly increased the fluorescence signal. This was very encouraging because if the membrane protein precipitated, it might pull down the sfGFP fused to it as well, which would result in lower fluorescence signal. If nanodiscs could solubilize the membrane protein-sfGFP fusion, we should get higher fluorescence, as most proteins were now in the supernatant.

To further test our hypothesis, we spun down all of the whole reaction mixes and took 1uL of each supernatant and ran both SDS-PAGE gels and western blots. In Supplementary Figure S9, the top panel is the western blot with anti-His antibody. Because the nanodiscs that we used have His tagged MSPs, we need to compare lanes 3-11 to lanes 13 and 14, which did not have DNA but nanodiscs only. Lanes 3, 6, and 9 had no significant bands in them, which was consistent with previous western blot results, suggesting that membrane protein would precipitate out when no detergent or nanodisc was present. Only when we added nanodiscs to the reaction mix could we see His tagged proteins (other than MSP) in the supernatant. All the non-MSP bands were labeled with red boxes in lane 4, 5, 7, 8, and 10. And when comparing these red boxes to those proteins in the second test, we can tell that these supernatant proteins were the same proteins from the whole reaction mix (same molecular weight), meaning that nanodiscs did help solubilize those membrane proteins. To further confirm those were the proteins of interest, we ran an exactly identical gel and performed the western blot with anti-GFP antibody. Results are shown in the middle

panel of Supplementary Figure S9. As we expected, the western blot results of the anti-GFP aligned well with the ones of the anti-His, suggesting that the bands marked with red boxes were our target proteins. As a comparison, the MSP (with His tag) bands only showed up in anti-His blot but not anti-GFP blot.

These are very encouraging results. We demonstrate that with the addition of nanodiscs, we can prevent TX-TL made membrane proteins from precipitating, and thus solubilize the membrane proteins in the supernatant. With the anti-GFP western blot, we show that those higher molecular weight bands can be the suspected dimers/oligomers of our target proteins.





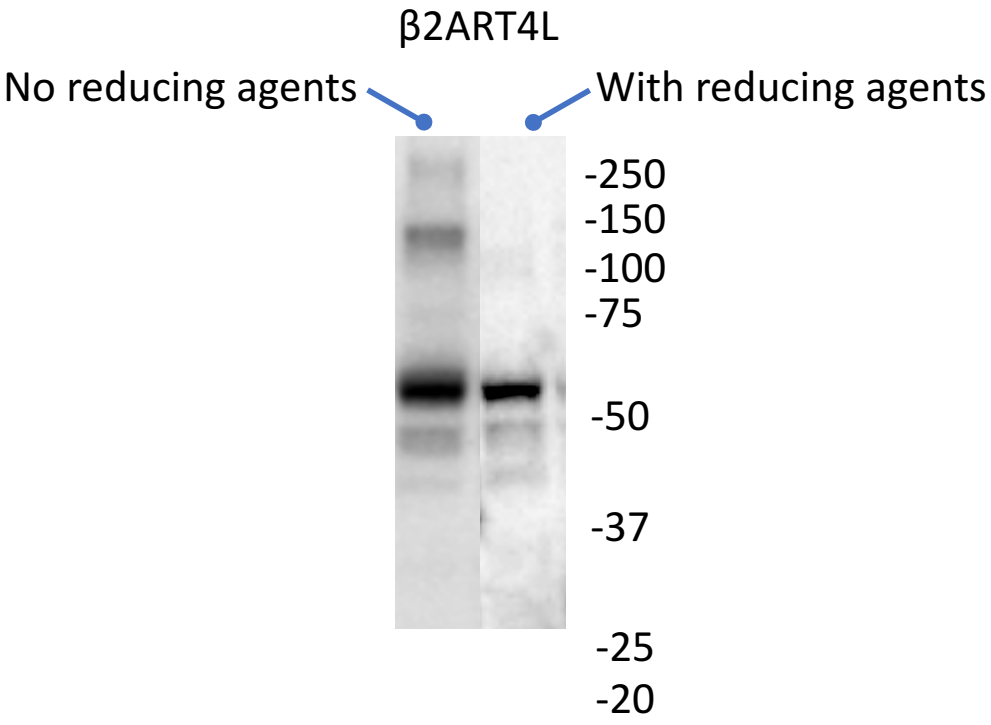
1	2	3	4	5	6
pSG73: b1ARts linear	pSG73: b1ARts linear w/ ND1	pSG73: b1ARts linear w/ ND3	pSG74: b2AR linear	pSG74: b2AR linear w/ ND1	pSG74: b2AR linear w/ ND3
7	8	9	10	11	12
pSG75: b2ART4L linear	pSG75: b2ART4L linear w/ ND1	pSG75: b2ART4L linear w/ ND3	Ladder	ND-MSP1D1-DMPC	ND-MSP1E3D1-DMPG

Supplementary Figure S9 Top: Western blot results of the TX-TL reaction samples with anti-His antibodies. Middle: Western blot results of the same gel with anti-GFP antibodies. Bottom: A table describing the samples in each lane.

4 A control experiment to verify the existence of dimers/oligomers

As we mentioned in the previous section, combining anti-His and anti-GFP western blots, we were confident that we observed not only GPCR-sfGFP monomers on the blots, but also their dimers/oligomers. All the gels were run without any reducing agents, thus in nondenaturing (native) conditions. To further verify the above assumption, we need to run a denaturing gel and western blot with reducing agents to see whether those higher molecular weight bands would disappear. If they would disappear in denaturing conditions, we would be certain that those bands were dimers/oligomers of GPCR-sfGFP.

TX-TL reactions (with nanodiscs), gels and western blots were performed using the same protocol, except when gel samples were prepared using supernatants from the TX-TL reactions, reducing agents and 10 min heating at 70°C were performed to create denatured proteins. Results are shown in Supplementary Figure S10. As we can see in the figure, there was possible existence of β2ART4L-sfGFP dimers between 100kD and 150kD at nondenaturing conditions. But the band disappeared in the lane with reducing agents, which can break the disulfide bonds that connect two monomers to be dimers. These results suggest that we have both monomers and dimers of the GPCR-sfGFP proteins.



Supplementary Figure S10 Western blot results of β2ART4L-sfGFP-His proteins with and without reducing agents.

SM2 Expressing a chimera potassium channel membrane protein KcsA-Kv1.3 in TX-TL

The first membrane protein that we tried to express and solubilize in TX-TL was a potassium channel membrane protein KcsA-Kv1.3, which was of interest to our collaborators at Amgen.

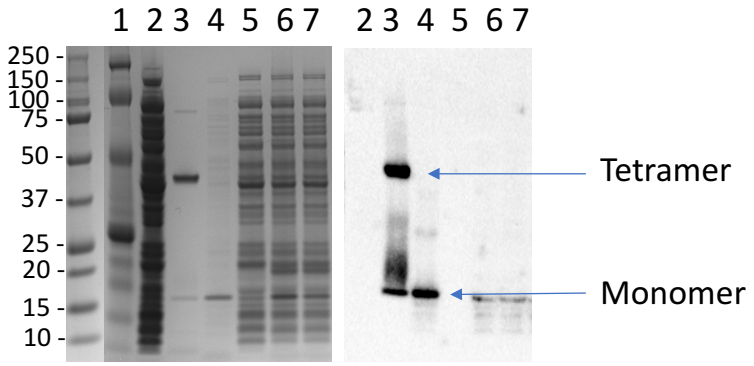
KcsA is a prokaryotic potassium channel from *Streptomyces lividans* [95] and Kv1.3 is a human voltage-gated potassium channel [96]. It has been shown that there is a remarkable structural conservation between the pore structures of the prokaryotic KcsA and eukaryotic Kv channels [97]. The chimeric KcsA-Kv1.3 is created by swapping subregions of the M1-M2 linker of KcsA with those of the S5-S6 linker of the Kv1.3. Because of that, not only is the chimeric potassium channel able to bind to specific toxins with high affinity, but it can also be expressed in *E. coli* [98].

As the TX-TL extract is based on *E. coli*, it would be reasonable to think that we might be able to express KcsA-Kv1.3 in TX-TL and solubilize it with membrane-like materials. However, Kv1.3 (we will use Kv1.3 instead of KcsA-Kv1.3 in the following paragraphs for simplicity) needs to form tetramers in order to be functional [99], which can be very challenging as there are few chaperone proteins in TX-TL to help the folding and formation of the tetramers.

To start, we tested a construct pCon-His^6-Kv1.3, in which the N-terminal His tagged Kv1.3 is constitutively expressed in TX-TL. We prepared six samples: 1) a negative control with only TX-TL reaction mix; 2) a positive control consisting of

160 ng purified Kv1.3 proteins in tetramer form; 3) the pellet from a reaction containing 15 nM Kv1.3 DNA (spun for 10 min at 14000 rpm to separate the pellet from the supernatant); 4) the supernatant from (3); 5) the sample from a TX-TL reaction containing 40 nM MSP1D1-His_DMPC (ND1); 6) the sample from a TX-TL reaction containing 40 nM MSP1E3D1-His_DMPG (ND3).

After incubating the TX-TL reactions at 29°C for at least 8 hours (except for the negative control), we took 1 μL out from each reaction mix and ran both SDS-PAGE gels and western blots. The results are shown in Supplementary Figure S10.



1	2	3	4	5	6	7
Ladder	TX-TL extract control	Purified Kv1.3 protein as positive control	Reaction without detergent (pellet)	Reaction without detergent (supernatant)	Reaction incubated with ND1	Reaction incubated with ND3

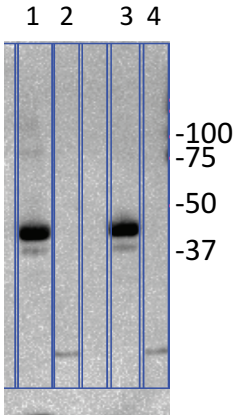
Supplementary Figure S11 Top left: SDS-PAGE gel results of the TX-TL reaction samples. Top right: Western blot results of the same gel. Bottom: A table describing the samples in each lane.

As we can see in the top left panel, the SDS-PAGE coomassie blue stain gel showed that 1) there was an abundance of proteins in the extract (lane 2); 2) purified Kv1.3 protein samples contained mostly tetramers and some monomers (lane 3); 3) the pellet sample in lane 4 showed the possible existence of Kv1.3 monomers made in TX-TL, and so did lanes 6 and 7 (among all the other protein bands from the extract). After analyzing the top right panel, which was the western blot with anti-His antibody, we confirmed the tetramers in the positive control (lane 3); in the same lane, we also saw some monomers and possible dimers among the smearing region. In lanes 4 and 5, we clearly observed that there was no Kv1.3 protein in the supernatant, but there were monomers and possibly dimers in the pellet. Most importantly, lane 6 and 7 showed that nanodiscs did not help the formation of Kv1.3 tetramers in TX-TL, which meant that it was very unlikely to have Kv1.3 proteins inserting into nanodiscs and forming tetramers spontaneously.

The construct we tested was pCon-His^6-Kv1.3, which had the His tag at the N terminus. There have been some studies on the differential effect of a His tag at the N- and C-termini [100]. So, we decided to make and test a construct pCon-Kv1.3-His^6 (pSG69), which had the His tag at the C terminus. Additionally, we would like to monitor the production of Kv1.3 in TX-TL and a convenient way to achieve this is fusing a fluorescent protein to Kv1.3. It has also been shown that some fluorescent proteins are prone to forming low-affinity oligomers and this tendency is boosted when fluorescent proteins are confined to membranes or when fused to naturally oligomeric proteins [101]. We thought this characteristic of fluorescent proteins

might help Kv1.3 monomers form tetramers when they were fused to fluorescent proteins *in vitro*. So, we designed and made a Kv1.3-sfGFP fusion construct pCon-Kv1.3-sfGFP-His^{6} (pSG66).

We tested the expression of both constructs in TX-TL with and without detergents. However, as the western blot shown in Supplementary Figure S11, we did not observe any hints of the formation of Kv1.3 tetramers, in neither C-terminal nor fusion constructs, after comparing lane 1 to lane 3 and lane 2 to lane 4. Not only did we test detergent Brij35 with pSG66 and pSG69, but we also tried these constructs with detergents DM (2X CMC), LMPG (5X CMC), CHAPS/DMPC&DMPG bicelles, and 12 nM and 40 nM nanodisc MSP1E3D1-His-DMPG, and none of them helped the formation of Kv1.3 tetramers (data not shown).



Supplementary Figure S12 The western blot results of pSG66 and pSG69 expressed in TX-TL with or without detergent Brij35. Lane 1 and 3 are pSG66 (Kv1.3-sfGFP fusion) samples; lane 1 has no detergent added and lane 3 has 10X CMC Brij35 added. Lane 2 and 4 are pSG69 (Kv1.3-His^{6}) samples; lane 2 has no detergent and lane 4 has 10X CMC Brij35. A ladder is added in lane 5. The blot was incubated with anti-His antibody.

After failing to generate Kv1.3 tetramers spontaneously in TX-TL with membrane-like materials, we changed our strategy to generate Kv1.3 monomers in

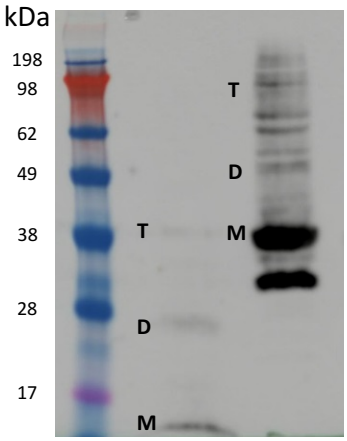
TX-TL and hoped to refold the misfolded Kv1.3 into correctly folded Kv1.3 tetramers.

Briefly, the refolding protocol is listed as below:

1. Set up 1 mL TX-TL reactions with corresponding DNA constructs (original construct, pSG66 or pSG69). After incubating the reactions at 29°C for at least 8 hours, spin down the reactions at 14,000 rpm for 10 min to harvest the pellets.
2. Incubate the TX-TL pellet with 200 μL 8M urea solution for 2 hours to unfold Kv1.3 proteins, and then purify the His tagged Kv1.3 proteins using HisPur Ni-NTA spin purification kit (including Ni-NTA resin on columns).
3. After incubating protein samples with the Ni-NTA resin for 2 hours, start refolding on columns with a pump circulating refolding buffer (20mM Tris pH7.5, 100 mM NaCl, 20 mM DM, 5 mM DMPG) and leave overnight. Then elute proteins with 500mM imidazole with refolding buffer.

After elution, refolded Kv1.3 proteins were used to run on SDS-PAGE and western blot. Supplementary Figure S12 shows the western blot results of the refolded proteins Kv1.3-His^6 and Kv1.3-sfGFP-His^6. As we can see, the middle lane shows that after refolding there were very likely traces of Kv1.3 monomers, dimers, and tetramers on the blot, though the amount of tetramers was quite limited. In the

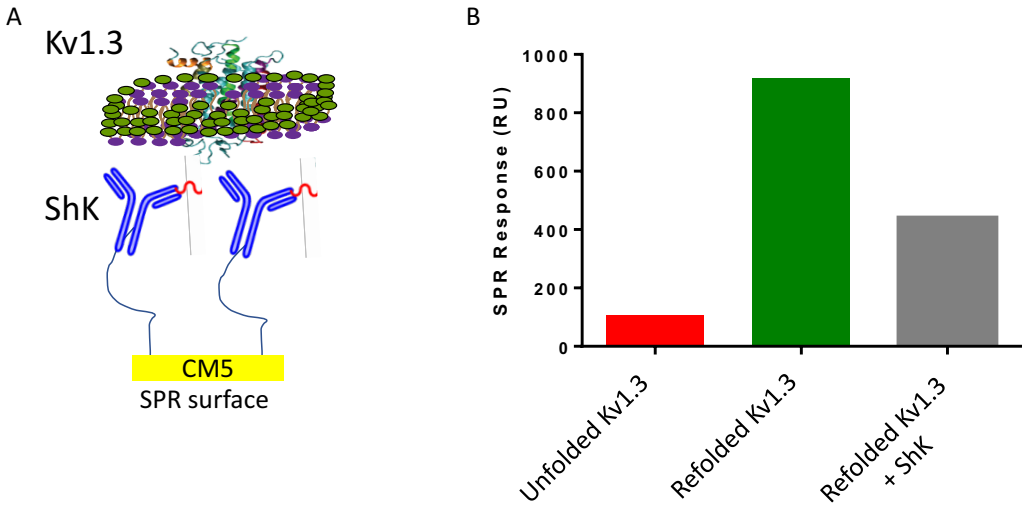
right lane, we can see a much higher yield of Kv1.3-sfGFP-His^6 proteins, which is consistent with our previous observation in Figure S11. We could have traceable amounts of both Kv1.3-sfGFP dimers and tetramers after the refolding process, and we also saw an intense band under the monomer band, which could be truncated sfGFP-His^6.



Supplementary Figure S13 The western blot results of refolded Kv1.3-His^6 and Kv1.3-sfGFP-His^6 proteins. A ladder is added in the left lane. The middle lane is the refolded Kv1.3-His^6. The right lane is the refolded Kv1.3-sfGFP-His^6. The blot was incubated with anti-His antibody.

The western blotting results were promising but not conclusive. If we have correctly folded tetramers, we could potentially verify their activity by testing their binding to the blocker ShK peptides. A 35-amino acid-residue polypeptide ShK from the sea anemone *Stichodactyla helianthus* has been shown to be a potent blocker for the Kv1.3 channel [102]. To perform the binding assay, we used a surface plasmon resonance (SPR)-based optical biosensor. We tethered ShK peptides to a CM5 chip and tested the binding of different Kv1.3 proteins by flowing samples through the individual channel on the chip (Supplementary Figure S13A). The results are shown in Supplementary Figure S13B. As we can see, unfolded Kv1.3 proteins were poorly bound to the ShK-tethered surface; in contrast, the refolded Kv1.3 proteins showed

significant binding response to the surface. Moreover, the refold Kv1.3 incubated with free ShK peptides showed significantly lower binding to the surface, indicating the binding between Kv1.3 and ShK-CM5 was specific.



Supplementary Figure S14 A: An illustrative diagram describes a binding assay designed for Kv1.3 proteins. The SPR chip was CM5 and was coated with ShK peptides. Active Kv1.3 proteins flowed through the chip can bind to the ShK peptides and result in an increase of the SPR response unit. B: A bar chart of the SPR binding assay results. The Y axis is the SPR response unit measured from a Biocore 2000. Unfolded Kv1.3 is the protein sample incubated in 8M urea solution for 2 hours; refolded Kv1.3 + ShK is the refolded Kv1.3 protein samples incubated with ShK peptides for 1 hour.

In conclusion, we failed to express active Kv1.3 membrane proteins in TX-TL with membrane-like materials, including nanodiscs; however, with an additional refolding process post-TX-TL, we were able to produce active Kv1.3 tetramers showing specific binding to their blocker, significantly advancing our understanding of the potential of the TX-TL platform.

Appendix I

Quantification of Terminator Strengths Using Linear DNAs in TX-TL

An important part of the DNA constructs, but sometimes overlooked, is the transcription terminator, which is a section of nucleic acid sequence that marks the end of a gene or operon in genomic DNA during transcription [103, 104]. Intrinsic transcription terminators of prokaryotes are distinguished by a common RNA motif – a stem loop structure [105]. A recent study on the transcription terminators characterizes 582 natural and synthetic terminators and their respective termination strength [106]. That study provides great insights into the understanding of terminators and useful guidance on choosing terminators when building synthetic biocircuits. In that study, terminator strength was determined from a simple assay that compares the expression of two fluorescent reporters, one placed before and one after the terminator, and the assay was done using plasmids in cells. Because we often use linear DNAs instead of plasmids when we prototype biocircuits in the cell-free transcription-translation system (TX-TL), we would like to find out whether the terminators would perform any differently if this were done using linear DNAs in the (TX-TL).

To that end, we chose a standard terminator (T500) that we used in our lab and four of the strongest terminators (T13-16) from [106]. We also included a T-neg, which is a short sequence with no significant secondary structure, as a negative control. We chose to use CFP as the main reporter fluorescent protein and GFP as the

leaky reporter. We started with DNA1, which has a pLac promoter activated by IPTG and a strong ribosome binding site (RBS) UTR1. All the DNA1s have CFP gene and one of the six terminators (Figure 1). We also built DNA2, consisting of only a RBS, a gene and a terminator; as a result of no promoter, DNA2 itself cannot be transcribed. We then used the Gibson assembly to stitch DNA1 with DNA2 and generated six different DNA1-2s. For DNA1s and DNA1-2s, they can be activated with the addition of IPTG and produce CFP regardless of what terminators are used. The CFP signal should also be the same among all the DNAs when activated with the same concentration of IPTG as they all share the same promoter, RBS and gene.

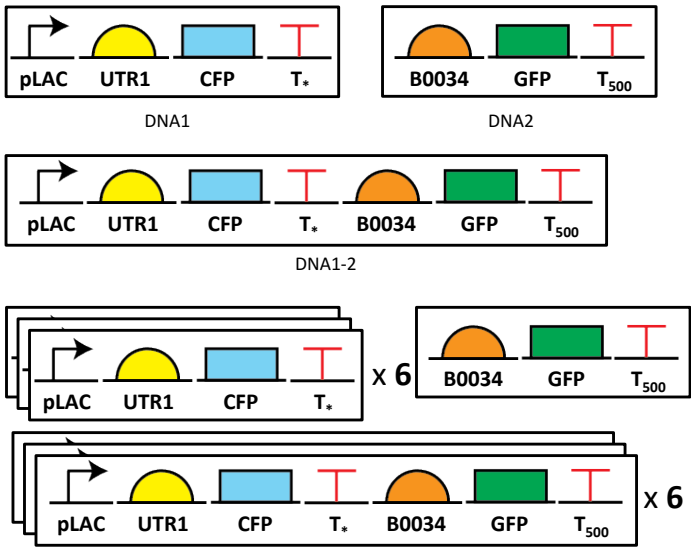


Figure 1 Illustration diagrams of constructs used for the terminator strength assay. Upper panel: DNA1: linear DNAs pLac-UTR1-CFP with the selected terminators T*, respectively. DNA2: a piece of linear DNA consisting of only RBS-B0034, reporter GFP and terminator T500 but no promoter. DNA1-2: linear DNAs pLac-UTR1-CFP-T*-B0034-GFP-T500, made by ligating DNA1 and DNA2 together using Gibson assembly. Bottom panel: the terminator strength assay was performed with six different DNA1, one DNA2, and six different DNA1-2, and each had three repeats.

DNA1-2s can also produce GFP fluorescence, which is an indicator of how strong the terminator T* is. The stronger the terminator T* is, the lower the GFP fluorescence will be generated (less leaky). We set up 14 TX-TL reactions (times 3 repeats) with six different DNA1s, six different DNA1-2s, a positive control, and a negative control (DNA2). After 12 hours, we measured the CFP and GFP fluorescences of all the samples and the normalized results were plotted in Figure 2. As we can see in Figure 2A, DNA2, the negative control, had no expression as expected. DNA1s had significant CFP signals but little to no GFP signals, also as expected. DNA1-2s showed varied results, depending on which terminators were used. The T-neg-B, which has no terminator between, showed the highest leaky GFP expression. And the T13-B with the strongest terminator T13 showed the lowest GFP signal.

With these encouraging results, we quantify the terminator strength using the following equation adapted from the one in [106]:

$$Terminator\ strength = \left(\frac{CFP_B}{GFP_B}\right)\left(\frac{CFP_0}{GFP_0}\right)^{-1},$$

in which CFP_B and GFP_B are the CFP and GFP fluorescent signals from DNA1-2s (T500-B, T13-16-B, T-neg-B in Figure 2), and CFP_0 and GFP_0 are the fluorescent signals from DNA1s (T500, T13-T16, T-neg in Figure 2). The terminator strength of each terminator was then calculated and plotted in Figure 2B. Comparing the results to the ones in [106] plotted in Figure 2C, they were consistent, in which T13 was the strongest and T14-T16 were similar but stronger than T500.

To summarize, we characterized and quantified the terminator strengths of the selected terminators using linear DNAs in TX-TL. We found that terminators in TX-TL behaved at similar relative strengths to *in vivo* experiments, although much lower efficiency. We suspect that linear DNAs are more flexible and have less rigid secondary structures compared to their circular plasmid counterparts. This could affect the rigidity of the stem-loop structure of the terminators, resulting in leakier transcription. While we would like to note that with linear DNAs in TX-TL, one does not need terminators after the coding sequences as we usually have only one coding sequence on one piece of linear DNA and leaky transcription should not affect the outcomes at all. But this study provides an opportunity to examine the terminators in different media and different environments, reminding us of the differences that we need to take into account when translating results from one platform to the other.

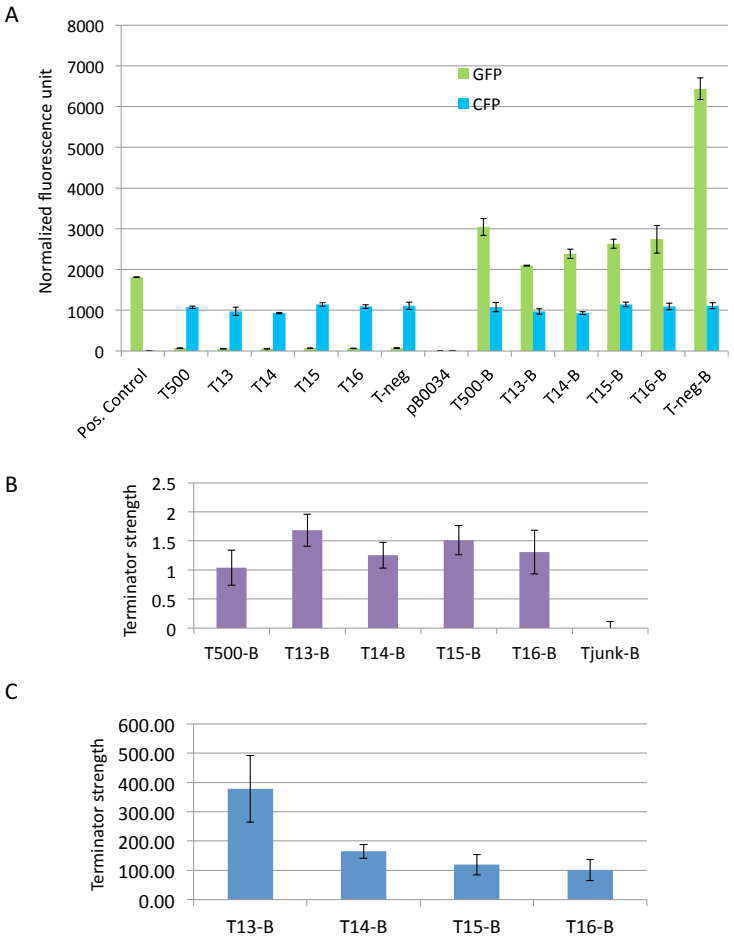


Figure 2 The terminator strength assay data. **A**: Normalized fluorescence data plotted in a bar chart. For every sample, two fluorescence measurements were taken (CFP: excitation 430 nm and emission 470 nm; GFP: excitation 485 nm and emission 525 nm) after 12-hour incubation of the TX-TL reactions. For clear comparisons, we normalized the CFP fluorescences from the samples with the same terminator (T500 and T500-B and etc.) to be equal while keeping CFP/GFP fluorescence ratio constant. The pos. control is a sample with pCon-GFP DNA as a positive control. Sample pB0034 is the DNA2. Samples T500, T13-T16 and T-neg are DNA1s, and T500-B, T13-16-B, and T-neg-B are DNA1-2s. **B**: A bar chart showing the normalized terminator strengths calculated from the data in A using the terminator strength equation. **C**: A bar chart showing the terminator strength data of the same terminators from [106].

CONCLUSION AND FUTURE DIRECTIONS

In this thesis, we investigate several biological circuits in a cell-free transcription-translation system (TX-TL). We demonstrated three distinct aspects of biological circuits: circuit dynamics (Chapter 1), phosphorylation (Chapter 2), and membrane proteins (Chapter 3). We successfully prototype a feedforward loop circuit using mathematical modeling in the TX-TL system and further implement it in cells. We also prototype a phosphorylation-based insulator circuit in the TX-TL and perform system identification of a mathematical model for the circuit by leveraging the advantages of the cell-free system. We further demonstrate that the TX-TL system could be used to prototype circuits involving membrane proteins by expressing biologically active membrane proteins in the system with extra engineering. Overall, these results bring us one step closer to the goals of utilizing biological circuits to control cells.

Prototyping biocircuits in the TX-TL system is the first step towards engineering biological systems in a fashion that resembles electrical engineering. A cellphone that is capable of doing many things is made of thousands of modular electronic components. If we were to one day build a cell with a specific function like assembling a cellphone, we need to gain much more understanding of the parts underlying the cell. If we were to one day develop novel, functional, and modular biological circuits in the same manner as building electronic circuits on a breadboard, we ought to prototype parts on a “biomolecular breadboard”, such as the TX-TL system.

In the TX-TL system, one can skip the cell growth step, and go straight into parts characterization. It has been shown (in this thesis and previous studies) that the TX-TL is a helpful platform for prototyping because of its simplicity, freedom, and convenience. We have described in the thesis some of the limitations of the current TX-TL system: limited resources, batch-to-batch variations, and *vitro*-to-*vivo* inconsistency. To make the TX-TL more instructive, we should supply the TX-TL reactions with continuous fresh resources. We should standardize the characterization process of the parts underlying biological circuits. We should develop a reference or an algorithm to offset the batch-to-batch variations by normalizing all major parameters towards the same standard. We should account for the differences between the cell-free system and cell-based system to resolve the inconsistency between them with a sufficient mathematical model.

Bibliography

1. Guo, S. and R.M. Murray, *Prototyping And Implementation Of A Novel Feedforward Loop In A Cell-Free Transcription-Translation System And Cells*. bioRxiv, 2017.
2. Guo, S., E. Yeung, and R.M. Murray, *Implementation and System Identification of a Phosphorylation-Based Insulator in a Cell-Free Transcription-Translation System*. bioRxiv, 2017.
3. Guo, S., A. Vaish, Q. Chen, and R.M. Murray, *Expressing Biologically Active Membrane Proteins in a Cell-Free Transcription-Translation Platform*. bioRxiv, 2017.
4. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators*. Nature, 2000. **403**(6767): p. 335-8.
5. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli*. Nature, 2000. **403**(6767): p. 339-42.
6. Jinek, M., K. Chylinski, I. Fonfara, et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity*. Science, 2012. **337**(6096): p. 816-21.
7. Cong, L., F.A. Ran, D. Cox, et al., *Multiplex genome engineering using CRISPR/Cas systems*. Science, 2013. **339**(6121): p. 819-23.
8. Nielsen, A.A., B.S. Der, J. Shin, et al., *Genetic circuit design automation*. Science, 2016. **352**(6281): p. aac7341.
9. Shendure, J., G.J. Porreca, N.B. Reppas, et al., *Accurate multiplex polony sequencing of an evolved bacterial genome*. Science, 2005. **309**(5741): p. 1728-32.
10. Margulies, M., M. Egholm, W.E. Altman, et al., *Genome sequencing in microfabricated high-density picolitre reactors*. Nature, 2005. **437**(7057): p. 376-80.
11. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age*. Nat Rev Genet, 2010. **11**(5): p. 367-79.
12. Purnick, P.E. and R. Weiss, *The second wave of synthetic biology: from modules to systems*. Nat Rev Mol Cell Biol, 2009. **10**(6): p. 410-22.
13. Hsiao, V., Y. Hori, P.W. Rothemund, and R.M. Murray, *A population-based temporal logic gate for timing and recording chemical events*. Mol Syst Biol, 2016. **12**(5): p. 869.
14. Hsiao, V., E.L. de los Santos, W.R. Whitaker, et al., *Design and implementation of a biomolecular concentration tracker*. ACS Synth Biol, 2015. **4**(2): p. 150-61.
15. Martin, V.J., D.J. Pitera, S.T. Withers, et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids*. Nat Biotechnol, 2003. **21**(7): p. 796-802.
16. Hale, V., J.D. Keasling, N. Renninger, and T.T. Diagana, *Microbially derived artemisinin: a biotechnology solution to the global problem of access to affordable antimalarial drugs*. Am J Trop Med Hyg, 2007. **77**(6 Suppl): p. 198-202.
17. Ro, D.K., E.M. Paradise, M. Ouellet, et al., *Production of the antimalarial drug precursor artemisinic acid in engineered yeast*. Nature, 2006. **440**(7086): p. 940-3.
18. Bagh, S., M. Mazumder, T. Velauthapillai, et al., *Plasmid-borne prokaryotic gene expression: sources of variability and quantitative system characterization*. Phys Rev E Stat Nonlin Soft Matter Phys, 2008. **77**(2 Pt 1): p. 021919.
19. Rosenfeld, N., J.W. Young, U. Alon, et al., *Gene regulation at the single-cell level*. Science, 2005. **307**(5717): p. 1962-5.
20. Shin, J. and V. Noireaux, *An E. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells*. ACS Synth Biol, 2012. **1**(1): p. 29-41.

21. Sun, Z.Z., C.A. Hayes, J. Shin, et al., *Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology.* J Vis Exp, 2013(79): p. e50762.
22. Sun, Z.Z., E. Yeung, C.A. Hayes, et al., *Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system.* ACS Synth Biol, 2014. **3**(6): p. 387-97.
23. Niederholtmeyer, H., Z.Z. Sun, Y. Hori, et al., *Rapid cell-free forward engineering of novel genetic ring oscillators.* Elife, 2015. **4**: p. e09771.
24. Takahashi, M.K., C.A. Hayes, J. Chappell, et al., *Characterizing and prototyping genetic networks with cell-free transcription-translation reactions.* Methods, 2015. **86**: p. 60-72.
25. Cardinale, S. and A.P. Arkin, *Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems.* Biotechnol J, 2012. **7**(7): p. 856-66.
26. Murphy, K.C., *Lambda Gam protein inhibits the helicase and chi-stimulated recombination activities of Escherichia coli RecBCD enzyme.* J Bacteriol, 1991. **173**(18): p. 5808-21.
27. Shen-Orr, S.S., R. Milo, S. Mangan, and U. Alon, *Network motifs in the transcriptional regulation network of Escherichia coli.* Nat Genet, 2002. **31**(1): p. 64-8.
28. Milo, R., S. Shen-Orr, S. Itzkovitz, et al., *Network motifs: simple building blocks of complex networks.* Science, 2002. **298**(5594): p. 824-7.
29. Mangan, S. and U. Alon, *Structure and function of the feed-forward loop network motif.* Proc Natl Acad Sci U S A, 2003. **100**(21): p. 11980-5.
30. Tuza, Z.A., V. Singhal, J. Kim, and R.M. Murray, *An In Silico Modeling Toolbox for Rapid Prototyping of Circuits in a Biomolecular "Breadboard" System.* 2013 Ieee 52nd Annual Conference on Decision and Control (Cdc), 2013: p. 1404-1410.
31. Sarrion-Perdigones, A., E.E. Falconi, S.I. Zandalinas, et al., *GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules.* PLoS One, 2011. **6**(7): p. e21622.
32. Lee, N., C. Francklyn, and E.P. Hamilton, *Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter.* Proc Natl Acad Sci U S A, 1987. **84**(24): p. 8814-8.
33. Ramos, J.L., M. Martinez-Bueno, A.J. Molina-Henares, et al., *The TetR family of transcriptional repressors.* Microbiol Mol Biol Rev, 2005. **69**(2): p. 326-56.
34. Li, X., X. Zhao, Y. Fang, et al., *Generation of destabilized green fluorescent protein as a transcription reporter.* J Biol Chem, 1998. **273**(52): p. 34970-5.
35. Gottesman, S., E. Roche, Y. Zhou, and R.T. Sauer, *The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system.* Genes Dev, 1998. **12**(9): p. 1338-47.
36. Wojtkowiak, D., C. Georgopoulos, and M. Zylicz, *Isolation and characterization of ClpX, a new ATP-dependent specificity component of the Clp protease of Escherichia coli.* J Biol Chem, 1993. **268**(30): p. 22609-17.
37. Siegal-Gaskins, D., Z.A. Tuza, J. Kim, et al., *Gene circuit performance characterization and resource usage in a cell-free "breadboard".* ACS Synth Biol, 2014. **3**(6): p. 416-25.
38. Nirenberg, M.W. and J.H. Matthaei, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides.* Proc Natl Acad Sci U S A, 1961. **47**: p. 1588-602.
39. Matthaei, H. and M.W. Nirenberg, *The dependence of cell-free protein synthesis in E. coli upon RNA prepared from ribosomes.* Biochem Biophys Res Commun, 1961. **4**: p. 404-8.
40. Carlson, E.D., R. Gan, C.E. Hodgman, and M.C. Jewett, *Cell-free protein synthesis: applications come of age.* Biotechnol Adv, 2012. **30**(5): p. 1185-94.

41. Davis, J.H., A.J. Rubin, and R.T. Sauer, *Design, construction and characterization of a set of insulated bacterial promoters.* Nucleic Acids Res, 2011. **39**(3): p. 1131-41.
42. Wang, P., L. Robert, J. Pelletier, et al., *Robust growth of Escherichia coli.* Curr Biol, 2010. **20**(12): p. 1099-103.
43. Del Vecchio, D., A.J. Ninfa, and E.D. Sontag, *Modular cell biology: retroactivity and insulation.* Mol Syst Biol, 2008. **4**: p. 161.
44. Jayanthi, S., K.S. Nilgiriwala, and D. Del Vecchio, *Retroactivity controls the temporal dynamics of gene transcription.* ACS Synth Biol, 2013. **2**(8): p. 431-41.
45. Nilgiriwala, K.S., J. Jimenez, P.M. Rivera, and D. Del Vecchio, *Synthetic tunable amplifying buffer circuit in E. coli.* ACS Synth Biol, 2015. **4**(5): p. 577-84.
46. Ninfa, A.J. and P. Jiang, *PII signal transduction proteins: sensors of alpha-ketoglutarate that regulate nitrogen metabolism.* Curr Opin Microbiol, 2005. **8**(2): p. 168-73.
47. Hurst, R., Ohana, R. and Nath, N., *Protein Interactions Analysis Using Cell-Free Protein Expression Systems and HaloLink™ Protein Arrays.* 2011.
48. Vajda, S., J.J. Distefano, 3rd, K.R. Godfrey, and J. Fagarasan, *Parameter space boundaries for unidentifiable compartmental models.* Math Biosci, 1989. **97**(1): p. 27-60.
49. Yeung, E., S. Guo, and R.M. Murray, *System identification of phosophorylation based insulator in a cell-free in vitro transcription-translation system.* Winter Q-Bio Conference, 2014.
50. Guo, S., E. Yeung, K.S. Nilgiriwala, et al., *Implementation And Simulation Of Phosphorylation-Based Insulator In Transcription-Translation Platform.* Winter Q-Bio Conference, 2014.
51. Pioszak, A.A. and A.J. Ninfa, *Genetic and biochemical analysis of phosphatase activity of Escherichia coli NRII (NtrB) and its regulation by the PII signal transduction protein.* J Bacteriol, 2003. **185**(4): p. 1299-315.
52. Ninfa, A.J., L.J. Reitzer, and B. Magasanik, *Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers.* Cell, 1987. **50**(7): p. 1039-46.
53. Ninfa, A.J. and B. Magasanik, *Covalent modification of the glnG product, NRI, by the glnL product, NRII, regulates the transcription of the glnALG operon in Escherichia coli.* Proc Natl Acad Sci U S A, 1986. **83**(16): p. 5909-13.
54. Vogel, U. and K.F. Jensen, *The RNA chain elongation rate in Escherichia coli depends on the growth rate.* J Bacteriol, 1994. **176**(10): p. 2807-13.
55. Paige, J.S., K.Y. Wu, and S.R. Jaffrey, *RNA mimics of green fluorescent protein.* Science, 2011. **333**(6042): p. 642-6.
56. Takahashi, M.K., J. Chappell, C.A. Hayes, et al., *Rapidly characterizing the fast dynamics of RNA genetic circuitry with cell-free transcription-translation (TX-TL) systems.* ACS Synth Biol, 2015. **4**(5): p. 503-15.
57. Capriotti, E. and R. Casadio, *K-Fold: a tool for the prediction of the protein folding kinetic order and rate.* Bioinformatics, 2007. **23**(3): p. 385-6.
58. Del Vecchio, D. and R.M. Murray, *Biomolecular feedback systems.* p. 1 online resource (286 pages).
59. Elf, J., G.W. Li, and X.S. Xie, *Probing transcription factor dynamics at the single-molecule level in a living cell.* Science, 2007. **316**(5828): p. 1191-4.
60. Pioszak, A.A. and A.J. Ninfa, *Mutations altering the N-terminal receiver domain of NRI (NtrC) That prevent dephosphorylation by the NRII-PII complex in Escherichia coli.* J Bacteriol, 2004. **186**(17): p. 5730-40.

61. Sniegowski, J.A., J.W. Lappe, H.N. Patel, et al., *Base catalysis of chromophore formation in Arg96 and Glu222 variants of green fluorescent protein.* J Biol Chem, 2005. **280**(28): p. 26248-55.
62. Shin, J. and V. Noireaux, *Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70.* J Biol Eng, 2010. **4**: p. 8.
63. Shaner, N.C., R.E. Campbell, P.A. Steinbach, et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat Biotechnol, 2004. **22**(12): p. 1567-72.
64. Jiang, P., A.C. Ventura, E.D. Sontag, et al., *Load-induced modulation of signal transduction networks.* Sci Signal, 2011. **4**(194): p. ra67.
65. Gyorgy, A. and D. Del Vecchio, *Retroactivity to the Input in Complex Gene Transcription Networks.* 2012 Ieee 51st Annual Conference on Decision and Control (Cdc), 2012: p. 3595-3601.
66. Ossareh, H.R., A.C. Ventura, S.D. Merajver, and D. Del Vecchio, *Long signaling cascades tend to attenuate retroactivity.* Biophys J, 2011. **100**(7): p. 1617-26.
67. Jayanthi, S. and D. Del Vecchio, *Retroactivity Attenuation in Bio-Molecular Systems Based on Timescale Separation.* Ieee Transactions on Automatic Control, 2011. **56**(4): p. 748-761.
68. Jayanthi, S. and D. Del Vecchio, *On the Compromise between Retroactivity Attenuation and Noise Amplification in Gene Regulatory Networks.* Proceedings of the 48th Ieee Conference on Decision and Control, 2009 Held Jointly with the 2009 28th Chinese Control Conference (Cdc/Ccc 2009), 2009: p. 4565-4571.
69. Shimizu, Y., A. Inoue, Y. Tomari, et al., *Cell-free translation reconstituted with purified components.* Nat Biotechnol, 2001. **19**(8): p. 751-5.
70. Cournia, Z., T.W. Allen, I. Andricioaei, et al., *Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory.* J Membr Biol, 2015. **248**(4): p. 611-40.
71. Choe, S., *Potassium channel structures.* Nat Rev Neurosci, 2002. **3**(2): p. 115-21.
72. Pierce, K.L., R.T. Premont, and R.J. Lefkowitz, *Seven-transmembrane receptors.* Nat Rev Mol Cell Biol, 2002. **3**(9): p. 639-50.
73. Bhate, M.P., K.S. Molnar, M. Goulian, and W.F. DeGrado, *Signal transduction in histidine kinases: insights from new structures.* Structure, 2015. **23**(6): p. 981-94.
74. Lappano, R. and M. Maggiolini, *G protein-coupled receptors: novel targets for drug discovery in cancer.* Nat Rev Drug Discov, 2011. **10**(1): p. 47-60.
75. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs.* FEBS Lett, 2010. **584**(9): p. 1721-7.
76. Ishihara, G., M. Goto, M. Saeki, et al., *Expression of G protein coupled receptors in a cell-free translational system using detergents and thioredoxin-fusion vectors.* Protein Expr Purif, 2005. **41**(1): p. 27-37.
77. Roos, C., L. Kai, D. Proverbio, et al., *Co-translational association of cell-free expressed membrane proteins with supplied lipid bilayers.* Mol Membr Biol, 2013. **30**(1): p. 75-89.
78. Roos, C., M. Zocher, D. Muller, et al., *Characterization of co-translationally formed nanodisc complexes with small multidrug transporters, proteorhodopsin and with the E. coli MraY translocase.* Biochim Biophys Acta, 2012. **1818**(12): p. 3098-106.
79. Rues, R.B., V. Dotsch, and F. Bernhard, *Co-translational formation and pharmacological characterization of beta1-adrenergic receptor/nanodisc complexes with different lipid environments.* Biochim Biophys Acta, 2016. **1858**(6): p. 1306-16.

80. Yang, J.P., T. Cirico, F. Katzen, et al., *Cell-free synthesis of a functional G protein-coupled receptor complexed with nanometer scale bilayer discs.* BMC Biotechnol, 2011. **11**: p. 57.
81. Schneider, B., F. Junge, V.A. Shirokov, et al., *Membrane protein expression in cell-free systems.* Methods Mol Biol, 2010. **601**: p. 165-86.
82. Christopher, J.A., J. Brown, A.S. Dore, et al., *Biophysical fragment screening of the beta1-adrenergic receptor: identification of high affinity arylpiperazine leads using structure-based drug design.* J Med Chem, 2013. **56**(9): p. 3446-55.
83. Cherezov, V., D.M. Rosenbaum, M.A. Hanson, et al., *High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor.* Science, 2007. **318**(5854): p. 1258-65.
84. Fredriksson, R., M.C. Lagerstrom, L.G. Lundin, and H.B. Schioth, *The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints.* Mol Pharmacol, 2003. **63**(6): p. 1256-72.
85. Lefkowitz, R.J., *Seven transmembrane receptors: something old, something new.* Acta Physiol (Oxf), 2007. **190**(1): p. 9-19.
86. Rosenbaum, D.M., S.G. Rasmussen, and B.K. Kobilka, *The structure and function of G-protein-coupled receptors.* Nature, 2009. **459**(7245): p. 356-63.
87. Terrillon, S. and M. Bouvier, *Roles of G-protein-coupled receptor dimerization.* EMBO Rep, 2004. **5**(1): p. 30-4.
88. Kasai, R.S. and A. Kusumi, *Single-molecule imaging revealed dynamic GPCR dimerization.* Curr Opin Cell Biol, 2014. **27**: p. 78-86.
89. Engler, C., R. Gruetzner, R. Kandzia, and S. Marillonnet, *Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes.* PLoS One, 2009. **4**(5): p. e5553.
90. Alexandrov, K. and W.A. Johnston, *Cell-free protein synthesis : methods and protocols.* Methods in molecular biology,. 2014, New York: Humana Press : Springer. xi, 313 p.
91. Nath, A., W.M. Atkins, and S.G. Sligar, *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins.* Biochemistry, 2007. **46**(8): p. 2059-69.
92. Angers, S., A. Salahpour, E. Joly, et al., *Detection of beta 2-adrenergic receptor dimerization in living cells using bioluminescence resonance energy transfer (BRET).* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3684-9.
93. Innis, R.B., F.M. Correa, and S.H. Synder, *Carazolol, an extremely potent beta-adrenergic blocker: binding to beta-receptors in brain membranes.* Life Sci, 1979. **24**(24): p. 2255-64.
94. Liu, J.J., R. Horst, V. Katritch, et al., *Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR.* Science, 2012. **335**(6072): p. 1106-10.
95. Schrempf, H., O. Schmidt, R. Kummerlen, et al., *A prokaryotic potassium ion channel with two predicted transmembrane segments from Streptomyces lividans.* EMBO J, 1995. **14**(21): p. 5170-8.
96. Cahalan, M.D. and K.G. Chandy, *Ion channels in the immune system as targets for immunosuppression.* Curr Opin Biotechnol, 1997. **8**(6): p. 749-56.
97. MacKinnon, R., S.L. Cohen, A. Kuo, et al., *Structural conservation in prokaryotic and eukaryotic potassium channels.* Science, 1998. **280**(5360): p. 106-9.

98. Legros, C., V. Pollmann, H.G. Knaus, et al., *Generating a high affinity scorpion toxin receptor in KcsA-Kv1.3 chimeric potassium channels.* J Biol Chem, 2000. **275**(22): p. 16918-24.
99. Spencer, R.H., Y. Sokolov, H. Li, et al., *Purification, visualization, and biophysical characterization of Kv1.3 tetramers.* J Biol Chem, 1997. **272**(4): p. 2389-95.
100. Mason, A.B., Q.Y. He, P.J. Halbrooks, et al., *Differential effect of a his tag at the N- and C-termini: functional studies with recombinant human serum transferrin.* Biochemistry, 2002. **41**(30): p. 9448-54.
101. Costantini, L.M., M. Fossati, M. Francolini, and E.L. Snapp, *Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions.* Traffic, 2012. **13**(5): p. 643-9.
102. Kalman, K., M.W. Pennington, M.D. Lanigan, et al., *ShK-Dap22, a potent Kv1.3-specific immunosuppressive polypeptide.* J Biol Chem, 1998. **273**(49): p. 32697-707.
103. Nudler, E. and I. Gusarov, *Analysis of the intrinsic transcription termination mechanism and its control.* Methods Enzymol, 2003. **371**: p. 369-82.
104. Gusarov, I. and E. Nudler, *The mechanism of intrinsic transcription termination.* Mol Cell, 1999. **3**(4): p. 495-504.
105. Cheng, S.W., E.C. Lynch, K.R. Leason, et al., *Functional importance of sequence in the stem-loop of a transcription terminator.* Science, 1991. **254**(5035): p. 1205-7.
106. Chen, Y.J., P. Liu, A.A. Nielsen, et al., *Characterization of 582 natural and synthetic terminators and quantification of their design constraints.* Nat Methods, 2013. **10**(7): p. 659-64.