DESIGN AND IMPLEMENTATION OF A GLUCOSE-RESPONSIVE GENETIC SWITCH CIRCUIT IN A CELL FREE SYSTEM

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Abhinav Adhikari

May 2021

ABSTRACT

Transcription factors (TF)—proteins that bind to promoter regions in the DNA to either activate or repress transcription—play an important role in biosensing and metabolic engineering applications. Metabolite-responsive TFs have the ability to sense specific metabolites; these small molecules bind to a specific domain on the TF and induce a conformational change of the protein, altering its DNA-binding strength and therefore the transcription rate of the downstream gene. We aim to implement, and further engineer, a prokaryote-derived TF, GntR, that responds to glucose or its derivative, D-gluconate, in a genetic switch circuit in a cell free protein synthesis (CFPS) system. Toward this aim, we constructed an *E. coli* promoter with the GntR operator site and performed preliminary tests of its repression and de-repression characteristics using CFPS. The CFPS platform has key advantages over in vivo systems, with particular regard to probing the metabolite, transcript, and protein levels due to the absence of a cell wall, and greatly facilitates validation of the metabolic and gene regulatory models we plan to develop for the circuit. Moreover, with recent key improvements to its reaction longevity and protein yield, CFPS is a promising platform for rapid prototyping and implementation of genetic circuits. Taken together, this work provides a framework for further optimization and development of the genetic switch circuit, which can be exploited for therapeutic and biosynthetic applications.

BIOGRAPHICAL SKETCH

Abhinav Adhikari was born in Kathmandu, Nepal. He attended Worcester Polytechnic Institute (WPI) and graduated in 2017 with a Bachelor of Science in Chemical Engineering. He received his Master of Science in Chemical Engineering from WPI in 2018 under the guidance of Professor Susan Roberts. In the fall of 2018, he joined the Robert Frederick Smith School of Chemical and Biomolecular Engineering at Cornell University and started his research with Dr. Jeffrey Varner.

This work is dedicated to my grandmother, Ranu Baral.

ACKNOWLEDGEMENTS

First of all, I would like to thank Professor Jeffrey Varner for his continued support and guidance in helping me become a better researcher. I would also like to thank Dr. Michael Vilkhovoy for mentoring and training me since I first joined the Varner Research Group. Thank you to my seniors and my friends in the Varner Group, especially Sandra Vadhin, Dr. Nicholas Horvath, Dr. David Dai, Dr. Rachel LeCover, Kritika Kashyap, Zhiping Zhang, Anirudh Murali, Abhishek Murti, Sruti Dammalapati, Zhihao Feng and Aasim Wani. Finally, I would like to thank my family for always being there for me.

TABLE OF CONTENTS

Biographical Sketch . . . iii
Dedication . . . iv
Acknowledgements . . . v
Table of Contents . . . vi
List of Tables . . . viii
List of Figures . . . ix

1 Introduction **1**
1.1 Cell free protein synthesis . . . 1
1.2 Origin and Preparation of Cell-Free Extracts . . . 3
1.3 Applications of Cell-Free Technologies . . . 7
1.3.1 Cell-Free Production of Biologics and Specialized Proteins 7
1.3.2 Cell-Free Systems in Synthetic Biology . . . 9
1.3.3 Cell-Free Metabolic Engineering . . . 11
1.4 Mathematical Modeling of Cell-Free Systems . . . 14
1.4.1 Cell-Free Transcription and Translation Models . . . 14
1.4.2 Metabolic Modeling Frameworks . . . 16
1.4.3 Emergence of Integrated Cell-Free Models . . . 19
1.5 Conclusions . . . 20
1.6 Acknowledgments . . . 23
1.7 Abbreviations . . . 24

2 Absolute Quantification of Cell-Free Protein Synthesis Metabolism by Reversed-Phase Liquid Chromatography-Mass Spectrometry **25**
2.1 Abstract . . . 25
2.2 Introduction . . . 26
2.3 Representative Results . . . 28
2.4 Materials and Methods . . . 29
2.5 Acknowledgements . . . 33

3 Effective Biophysical Modeling of Cell Free Transcription and Translation Processes **36**
3.1 Abstract . . . 36
3.2 Introduction . . . 37
3.3 Materials and Methods . . . 41
3.3.1 Cell free protein synthesis reactions. . . . 41
3.3.2 mRNA and protein quantification. . . . 42
3.3.3 Synthetic circuit architecture . . . 43
3.3.4 Formulation and solution of model equations . . . 44
3.3.5 Estimation of model parameters . . . 49
3.3.6 Morris sensitivity analysis . . . 52
3.4 Results . . . 53

3.4.1 Modeling and analysis of the $C1$ circuit 53
3.4.2 Modeling and analysis of the $C2$ circuit 56
3.5 Discussion 60
3.6 Supplementary Information for Effective Biophysical Modeling of Cell Free Transcription and Translation Processes 70
3.7 Acknowledgement 75

4 Developing a gluconate-responsive *E. coli* transcriptional element in an in vitro system **76**
4.1 Introduction 76
4.2 Materials and Methods 78
4.2.1 Construction of a modified bacterial promoter regulated by a gluconate-responsive repressor 78
4.2.2 Cell free synthesis in myTXTL extract 80
4.2.3 Cell free synthesis in PURE extract 80
4.2.4 Measurement of protein levels 81
4.2.5 Measurement of D-gluconate levels in the cell free extract 81
4.3 Results and Discussion 82
4.4 Conclusion 89

5 Conclusion and Future Directions **92**
5.1 Future Directions 93
5.1.1 Error-Prone PCR-Based Mutagenesis of GntR repressor protein and in vivo screening 93
5.1.2 Develop and characterize a glucose-responsive genetic switch in an *E. coli* cell free system 95
5.1.3 Computational modeling of transcription and translation (TX/TL) with integrated metabolic flux estimation. 97

Bibliography **98**

LIST OF TABLES

2.1 Each compound's corresponding limit of detection, range of linearity and correlation coefficient identified from standard curves. 34
2.2 Each compound's corresponding peak number, retention time, m/z value for 12C, 13C, and unlabeled, cone voltage, and MS species. 35

3.1 Characteristic parameters for TX/TL model equations. Key to references used in the table: (a) [59], (b) [227], (c) set by experiment, (d) [106], (e) estimated in this study, (f) [153], (g) [70], (h) calculated from plasmid sequence, (i) [136], (f) [237] 66
3.2 Estimated parameter values for the P70-deGFP model ($C1$). The mean and standard deviation of each parameter value was calculated over the ensemble of parameter sets meeting the rank selection criteria (N = 139). 67
3.3 Estimated parameter values for the negative feedback circuit ($C2$). The mean and standard deviation for each parameter was calculated over the ensemble of parameter sets (N = 498). 68

LIST OF FIGURES

1.1 Time line for milestones in the evolution of cell-free systems since 1950 until now. Abbreviations: cell-free protein synthesis (CFPS), G-coupled protein receptor (GCPR), transcription and translation (TXTL), CFPS system energized by PEP (PANOx), DNA-based Assembly and Synthesis of Hierarchical (DASH), commercially available *E. coli* cell free extract (myTXTL, Arbor Biosciences). 4

1.2 Schematic of cell-free protein synthesis. Cell extract is prepared by cell lysis, and cellular debris and chromosome DNA is removed. An energy source along with the necessary amino acids, nucleotides, and cofactor are added to the cell-free reaction. Template DNA of the target protein is added. The target protein is then easily purified from the cell-free system. Alternatively, cell-free extract can be freeze dried into pellets and paired with lyophilized DNA. Through the simple addition of water, proteins can be manufactured on site and on demand. Figure adapted from [25, 167]. 7

1.3 Schematic of the integration of transcription and translation processes with metabolism. Transcription and translation processes demand macromolecular precursors (e.g., NTPs, amino acids, and cofactors) from metabolism for gene expression. The integrated framework is represented as a stoichiometric matrix of metabolites participating in certain reactions, along with a description of the metabolic demands for protein expression. The metabolic flux is estimated subject to constraints, a pseudo-steady-state assumption, and an objective function. 20

2.1 Schematic of workflow for aniline tagging. The cell-free protein synthesis reaction is de-proteinized and tagged with 12C-aniline, while a standard stock mixture is tagged with 13C-aniline. Both mixtures are then mixed at a 1:1 volumetric ratio and analyzed by LC/MS. 30

2.2 Overlapped selected ion chromatograms for 40 metabolites. Mass chromatogram from a single LC/MS run of a 40μM standard mixture of 40 metabolites. Peaks were identified by their retention time and m/z values for each compound. Complete compound names and their abbreviations are listed in Table 2.1. 31

3.1 Schematic of the cell free gene expression circuits used in this study. **A**: Sigma factor 70 (σ_{70}) induced expression of deGFP. **B**: The circuit components encode for a negative feedback loop motif. Sigma factor 28 and deGFP-ssrA genes on the P70a promoters are expressed first because of the endogenous presence of sigma 70 factor in the extract. Sigma factor 28, once expressed, induces the P28a promoter, turning on the expression of the cI-ssrA gene which represses the P70a promoter. The circuit is modified from a previous study [59] by including an ssrA degradation tag on the cI gene. 44

3.2 Model simulations versus experimental measurements for σ_{70} induced deGFP expression. **A**: Simulated and measured deGFP mRNA concentration versus time using the small circuit G_{20} ensemble (N = 140). **B**: Simulated and measured deGFP protein concentration versus time the small circuit G_{20} ensemble (N = 140). **C**: Global sensitivity analysis of the P70-deGFP circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble. Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 140 parameter sets that were rank two or below. The parameter ensemble was used to calculate the mean (dashed line) and the 95% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates. 55

3.3 Model simulations versus experimental measurements for the negative feedback deGFP-ssrA circuit. **A**: Model simulations of the negative feedback deGFP-ssrA circuit using the G_{20} ensemble (N = 498). Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 489 parameter sets (rank two or below). The parameter ensemble was used to calculate the mean (dashed line) and the 99% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are also shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates. **B**: Global sensitivity analysis of the negative feedback deGFP-ssrA circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble. 57
3.4 Schematic of transcriptional (TX) and translation (TL) processes. 70

4.1 D-gluconate quantification. Glucose pulse and gluconate consumption experiments in myTXTL cell free extract. Glucose with a concentration approximating that of a typical diabetic patient (160 mg/dl) was added to the cell free reaction mixture at t = 1.25 hours. D-gluconate concentration, a downstream degradation product, was measured by LCMS with and without the addition of Glucose oxidase to the cell free mixture. Points denote the mean while the error bars denote one standard deviation . . 83
4.2 deGFP yield using 10 nM P70-deGFP linear DNA in myTXTL extract. The reaction was supplemented with 5μM GamS nuclease inhibitor. Shaded region denotes one standard deviation. 84
4.3 Repression of deGFP expression by *E. coli* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (2.5 nM) Error bars denote one standard deviation. 85
4.4 Repression of deGFP expression by *E. coli* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (2.5 nM) Error bars denote one standard deviation. 86
4.5 Repression of deGFP expression by *B. subtilis* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (5 nM) Error bars denote one standard deviation. 87
4.6 Repression of Venus expression by *E. coli* GntR in PURE system. Error bars denote one standard deviation. 90
4.7 De-repression of Venus expression by *E. coli* D-gluconate in PURE system. 10mM D-gluconate was added to the reaction mixture at time 0. Error bars denote standard error. 91

5.1 Map of plasmids used for the screening and selection procedure. *E. coli* cells used for the procedure will be doubly transformed with these plasmids. mP70 is the modified P70 promoter with added GntR operator site. 94
5.2 Schematic of the glucose sensor for Aim 1. mP70 is the modified P70 promoter with added GntR operator site. αS_{28} is the anti-sigma factor 28 that inhibits the P28 promoter, thereby repressing the expression of tdTomato. In the presence of high levels of gluconate, deGFP protein will be expressed along with αS_{28}. In low gluconate levels, only tdTomato will be expressed. 96

CHAPTER 1

INTRODUCTION

1.1 Cell free protein synthesis

[1] Cell-free biology is an emerging technology for research and the point-of-care manufacturing of a wide array of macromolecular and small molecule products. A distinctive feature of cell-free systems is the absence of cellular growth and maintenance, thereby allowing the direct allocation of carbon and energy resources toward a product of interest. Moreover, cell-free systems are more amenable than living systems to observation and manipulation, hence allowing rapid tuning of the reaction conditions. Recent advances in cell-free extract preparation and energy regeneration mechanisms have increased the versatility and range of applications that can be produced. Thus, the cell-free platform has evolved from merely an investigative research tool to a promising alternative to traditionally used living systems for biomanufacturing, as well as biological research. In combination with the rise of synthetic biology, cell-free systems today have not only taken on a new role as a promising technology for just-in-time manufacturing of therapeutically important biologics and high-value small molecules, but have also been utilized for applications such as biosensing, prototyping genetic parts, and metabolic engineering. They have also been used as educational tools at the high school and undergraduate levels for understanding synthetic biology due to their ease of use, rapid response times, and the availability of commercial kits for different cell-free platforms including

[1] Adapted with permission from Vilkhovoy, M.; Adhikari, A.; Vadhin, S.; Varner, J. D. The Evolution of Cell Free Biomanufacturing. Processes 2020, 8 (6), 675. https://doi.org/10.3390/pr8060675.

Escherichia coli, Chinese hamster ovary (CHO), HeLa, and plant cells [5, 33, 69]. Thus, cell-free technologies are promising tools that will likely be at the center of many future synthetic biology applications.

Arguably, today, the most widely used cell-free technology is cell-free protein synthesis (CFPS), an in vitro platform for protein transcription (TX) and translation (TL). The role of CFPS in research is not new (Figure 1.1); cell-free systems have been used for decades to explore fundamental biological mechanisms. For example, some of the first uses of CFPS were in the 1950s by Borsook [18] and Winnick [244], who used animal tissue homogenates to study how amino acids were incorporated into proteins. A few years later, *Staphylococcus aureus* extracts were used to confirm amino acid incorporation [58]. In 1956, the role of adenosine triphosphate (ATP) in protein production was discovered using rat liver extracts [82], and Nirenberg and Matthaei [132, 157] demonstrated templated translation, i.e., the now familiar codon code, using *E. coli* cell-free extracts (this work later led to a Nobel Prize in 1968). What arguably could be recognized as the first precursor to modern cell-free transcription and translation applications was developed in 1967 by Lederman and Zubay [114]; they developed a coupled transcription-translation bacterial extract that allowed DNA to be used as a template. Shortly after, Spirin and coworkers improved cell-free extract protein production with a continuous exchange of reactants and products, allowing the system to run for tens of hours; however, these systems could only synthesize a single product and were energy limited [203]. More recently, the energy efficiency of *E. coli* CFPS was improved by generating ATP with substrate level phosphorylation [109] and oxidative phosphorylation in the Cytomim system [98, 99, 96]. The use of glucose as an energy source was also explored [23]. Since oxidative phosphorylation is

a membrane associated process, the study of Swartz and colleagues revealed that membrane dependent energy metabolism can be activated in a cell-free system, suggesting complex metabolism is still occurring. Another platform, myTXTL [59], uses a different metabolic process that couples ATP regeneration and inorganic phosphate recycling to extend the duration of protein production. Synthetic genetic circuitry can also be constructed to control gene expression using a variety of approaches. Bacteriophage RNA polymerases are commonly used in CFPS for transcription. However, the use of a vast array of bacterial regulatory elements based on the sigma factor family has recently been explored, allowing multi-layer genetic cascades to be easily implemented [190, 192, 59].

Thus, the developments in CFPS have expanded its repertoire of applications, enabling it to be a viable alternative to living systems for not only investigative research, but also bioengineering and biomanufacturing at both small and large scales.

In this review, we discuss the evolution of cell-free technologies, in particular advancements in extract preparation, cell-free protein synthesis, and cell-free metabolic engineering applications. We then conclude with a discussion of the mathematical modeling of cell-free systems. Mathematical modeling of cell-free processes could be critical to addressing performance bottlenecks and estimating the costs of cell-free manufactured products.

1.2 Origin and Preparation of Cell-Free Extracts

There are two broad classes of cell-free systems: crude cell lysates and reconstituted systems. While crude extract-based systems, commonly prepared from

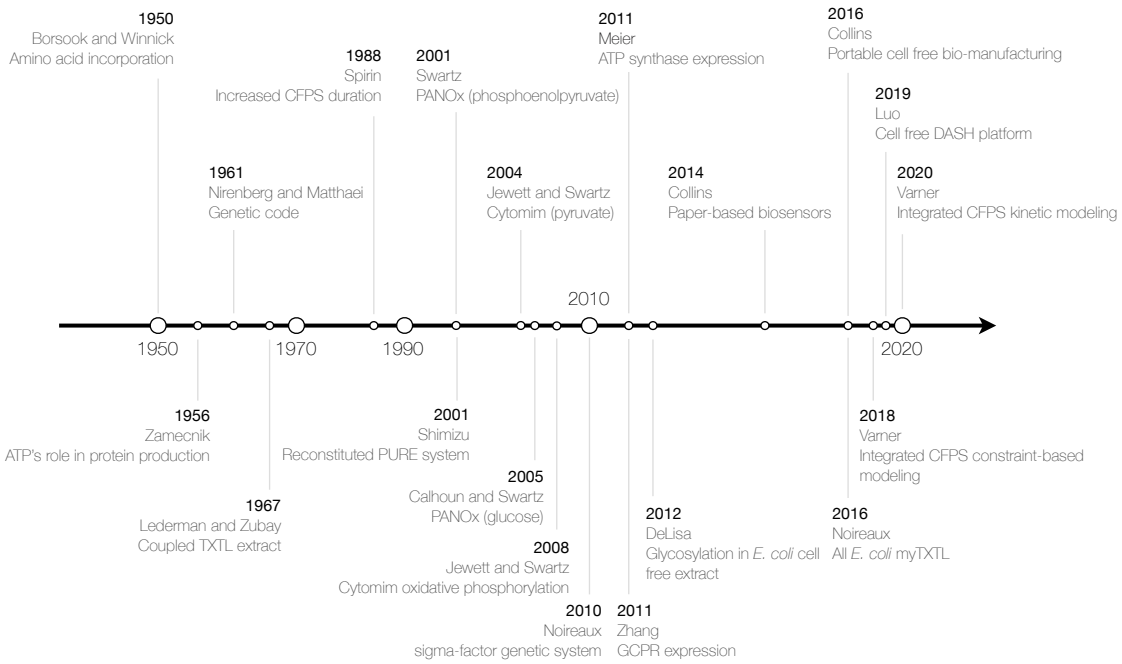


Figure 1.1: Time line for milestones in the evolution of cell-free systems since 1950 until now. Abbreviations: cell-free protein synthesis (CFPS), G-coupled protein receptor (GCPR), transcription and translation (TXTL), CFPS system energized by PEP (PANOx), DNA-based Assembly and Synthesis of Hierarchical (DASH), commercially available *E. coli* cell free extract (myTXTL, Arbor Biosciences).

E. coli, *S. cerevisiae*, rabbit reticulocytes, wheat germ, and insect cells [181], consist of the biocatalysts remaining after cell lysis, reconstituted systems are well defined, prepared using only the factors essential for protein synthesis: purified enzymes, tRNAs, ribosomes, amino acids, and energy molecules. The first purified extract of this kind, the Protein synthesis Using Recombinant Elements (PURE) system, was developed by Shimizu et al. in 2001 [189]. A similar system, based on *Thermus thermophilus*, was later developed by Zhou et al. [261]. Other specialized systems based on PURE have also been developed [93, 158, 148, 64]. These specialized systems have been utilized in applications including the synthesis of disulfide-bonded functional aglycosylated immunoglobulin G (IgG) and G-protein coupled receptors and the study of the effects of liposomes on

the solubility of aggregation-prone membrane proteins. Such reconstituted systems offer two main advantages over crude extracts. First, they are a valuable research tool for studying biological processes including protein expression and folding in the context of a completely defined reaction mixture. For example, given the precise knowledge and control of the components in the reconstituted system, it is possible to study the role of individual additions such as chaperones, translation elongation factors, ribosome release factors, and other molecules. Li et al., in a study analyzing the influence of such additions in the PURE system, showed that the efficiency of protein synthesis was limited by translation elongation capacity, ribosome release, and ribosome recycling. When the authors changed the ratio of elongation factors, release factors, and recycling factors to ribosome concentration to more closely resemble in vivo conditions, a five-fold improvement in the yield of firefly luciferase reporter protein was observed [118]. The second advantage of reconstituted systems, such as the PURE system, is that they do not contain proteases and nucleases, further improving the production of many proteins [210]. Despite these advantages, reconstituted systems suffer from two major drawbacks: a higher cost ($0.99/L for a PURE reaction vs. $0.15–0.57/L for a crude extract reaction; price for commercial kits [69]) and typically lower yields [81]. In this regard, crude cell extracts prevail; they are less expensive, especially for reactions carried out at larger scales [210]. They also offer more complex metabolic capabilities that can be exploited for energy regeneration, extending the duration of protein synthesis. Toward these advantages, the preparation of crude cell-free extracts, which has undergone a significant evolution since the early applications in the 1950s and 1960s, is now an area of considerable focus.

Cell-free extracts are commonly derived from crude cell extracts, where

the cell's transcription and translation machinery is retained while cellular debris and chromosomal DNA are discarded (Figure 1.2). Cells are typically grown until they reach the exponential phase, when they are harvested and lysed commonly using a high pressure homogenizer or a specialized bead mill [194]. Early extracts were prepared by centrifugation of lysates at 30,000× *g* followed by the addition of a mixture of amino acids, adenosine triphosphate (ATP), and other energy molecules, salts, and buffer [263, 174]. In the early 2000s, several changes were made by different research groups to make the extract preparation protocol more efficient including centrifugation at a lower rate (12,000× *g*), the use of shake flask fermentation, and the overexpression of the T7 RNA polymerase in the commercial BL21 (DE3) *E. coli* strain during extract preparation [108, 119, 110]. Alternatives to the high pressure homogenization step, which include the use of bead vortex mixing [194] or lytic enzymes [42], have also been recently explored. It is also possible to delete or overexpress certain genes in the source cell to yield customized cell-free extracts. For example, the Swartz group made several gene deletions in *E. coli* A19 cells before harvesting and extract preparation, improving the protein yield in the extract by up to 250% [24, 139]. More recently, a new extract design scheme was implemented by Jewett lab to mix different extracts combinatorially, each containing a unique overexpressed enzyme, to construct a full biosynthetic pathway [47]. The same group also optimized the extract preparation procedure to better accommodate the use of genetic circuits [197]. These developments showed that the extract preparation process can be modified depending on the end goal. However, there are still important unanswered questions in extract preparation. For instance, it still remains to be explored how one can selectively delete enzymes only in cell-free extracts. Continued research in this area could pave the way for

minimal extracts highly optimized for a known application.

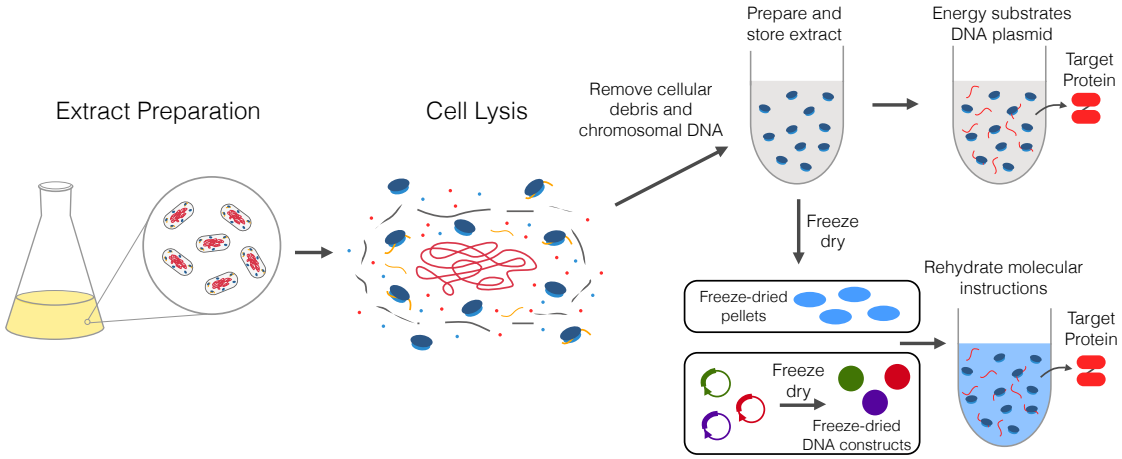


Figure 1.2: Schematic of cell-free protein synthesis. Cell extract is prepared by cell lysis, and cellular debris and chromosome DNA is removed. An energy source along with the necessary amino acids, nucleotides, and cofactor are added to the cell-free reaction. Template DNA of the target protein is added. The target protein is then easily purified from the cell-free system. Alternatively, cell-free extract can be freeze dried into pellets and paired with lyophilized DNA. Through the simple addition of water, proteins can be manufactured on site and on demand. Figure adapted from [25, 167].

1.3 Applications of Cell-Free Technologies

1.3.1 Cell-Free Production of Biologics and Specialized Proteins

Cell-free protein synthesis (CFPS) has been utilized in a wide range of applications from the expression of pharmaceutical proteins [122, 66, 151] to the production of libraries for protein evolution and structural genomics [213]. Complex post-translational modifications that are typically difficult to carry out using bacterial extracts have also recently been achieved cell-free. For example, N-linked glycoproteins have been produced in an *E. coli*-based cell-free extract

with the addition of a purified oligosaccharyltransferase (OST) and its lipid-linked oligosaccharides (LLOs) [71]. A single-pot glycoprotein synthesis system was also developed, potentially allowing for the production of personalized protein therapeutics [95]; in this system, the OSTs and LLOs were expressed in the *E. coli* host strain, which resulted in glycosylation-competent lysates. Other specialized proteins have also been produced in cell-free systems. The ability to add membrane mimics such as surfactants or liposomes to the extract as stability agents has allowed the production of membrane proteins [171, 133, 240, 193]. Vaccines [204, 151], protein assemblies [97, 56, 46], and proteins incorporating non-natural amino acids [7, 6, 129] have also been synthesized using CFPS systems. Disulfide-bonded proteins and antibodies have also been synthesized by adding components that facilitate the formation of these bonds in the mixture such as glutathione reductase, thioredoxin reductase, iodoacetamide, and disulfide isomerase (DsbC) [210, 175, 253, 148]. Cytosol-penetrating antibodies have also been synthesized [141]. The cell-free synthesis of onconase, a cytotoxic cancer therapeutic, has also been demonstrated [184]. Point-of-care protein manufacturing is also possible when microfluidic reactors are used. Compared to batch reactions, continuous flow microreactors typically offer users more precise control over mixing [78, 137]. An automated on-chip CFPS reactor has been developed that runs transcription and translation reactions simultaneously, but in separate compartments [63]. Each process can be optimized independently, and the quasi-continuous supply of new mRNA from the TX chamber to the TL chamber allows for a longer CFPS reaction and increased protein yield. Other microfluidic platforms integrate purification methods like dialysis and affinity chromatography [152, 149]. For example, cecropin B, an antimicrobial peptide that is widely used to control

biofilm-associated diseases, has been produced at a clinically-relevant dose in a few hours using a microfluidic device with on-chip protein purification [149]. A continuous exchange microfluidic reactor using a nanofabricated membrane to allow for extended reaction times and improved yields has also been developed with the goal of producing single-dose therapeutic proteins at the point-of-care [223]. Moreover, cell-free extracts can be lyophilized and stored at −80 °C for more than a year without degradation [210]. They can then be rehydrated with water and then incubated using the body's heat to activate the extract components, highlighting the portability and versatility of the cell-free platform [196].

Lastly, while the majority of CFPS has been carried out in small scales, there have been continued efforts to scale-up the technology. For example, the scalability of cell-free systems has been explored in a few academic studies [238, 212, 258, 161]. Moreover, there are also a few examples of industrial-scale cell-free protein production for high-value products such as antibodies and cytokines [258, 252] as well as industrial implementations of the technology, most notably by the clinical stage company Sutro Biopharma and the GreenWorX platform of GreenLight Biosciences. Continued research in the area of scale-up will potentially accelerate industrial adaptation.

1.3.2 Cell-Free Systems in Synthetic Biology

Applications of cell-free systems in synthetic biology are varied, from diagnostics to fundamental discovery and prototyping. Biosensing is an area where cell-free systems have recently proven useful. They possess a unique advan-

tage over whole cells because of their ability to detect species that are cytotoxic or impermeable to the cell wall [196]. These systems have been deployed to detect pathogens such as norovirus [125], Ebola virus [165], and Zika virus [166]. In addition, initial studies have shown that paper-based cell-free sensors can detect the presence of heavy metals such as mercury and drugs such as γ-hydroxybutyrate, by utilizing the transcriptional regulators, MerR and BlcR, respectively [68]. The portability offered by these systems further underscores their usefulness in the field. CFPS has also been used in the development of minimal cells, the simplest cellular mimics that consist of only the genes essential for survival. Minimal cells are often described as biological analogs to the hydrogen atom, which has served to uncover many fundamental phenomena in chemistry [146]. Their bottom-up construction has been made possible mainly by the advancements in two areas: compartmentalization strategies and programmable genetic circuits [62, 59, 92, 159, 105]. Toward this goal, proteins have been expressed in compartments such as liposomes [170, 59], phospholipid vesicles [192, 159], and hydrogel particles [260], and genetic circuits that encode oscillations [218, 153], negative feedback loops [100], or riboswitches acting as regulatory elements [130] have been developed in CFPS systems. Interacting minimal cells have also been developed [3]. Recently, the construction of dynamic biomaterials powered by artificial metabolism was also demonstrated by Hamada and coworkers, with applications ranging from locomotion, pathogen detection, and hybrid nanomaterials [76]. Cell free systems have also been used for prototyping novel genetic parts or circuits before using them in vivo. Prototyping can be done more efficiently in cell-free systems because of the tighter control over plasmid dosage, inducer concentrations, pH, temperature, and salt concentrations [196]. The ability to use linear PCR templates in

cell-free systems further accelerates this process [196]. Moreover, developments in experimental setup and analysis techniques such as the use of acoustic liquid-handling robots [145], real-time fluorescent reporters [154], microfluidics [40], and droplet-based expression [183] have allowed the prototyping to be carried out at high-throughput rates [196]. The relative ease of manipulating cell-free systems makes them attractive tools for investigating complex processes.

1.3.3 Cell-Free Metabolic Engineering

Cell-free systems have gained wide interest in metabolic engineering applications, primarily to circumvent the significant barriers of traditional in vivo processes [72]. For example, a major challenge in in vivo metabolic engineering is achieving high flux through synthetic pathways of interest. This is because cells have their own objectives, e.g., growth or maintenance, which drives metabolic flux away from the desired pathways. The complexity of living cells also makes computational modeling and optimization of metabolic flux difficult [48]. Cell-free systems, on the other hand, can be accurately modeled and the reaction environment tuned according to the bio-synthetic needs. They offer many advantages for the study, manipulation, and modeling of metabolism. Central amongst these is direct access to metabolites and the biosynthetic machinery without the interference of a cell wall. This allows interrogation of the chemical microenvironment, while the biosynthetic machinery is operating, potentially at a fine time resolution. Eliminating the need to maintain cell viability also allows the full allocation of energy resources to the production of products of interest. Because of these benefits, various metabolic engineering endeavors have been made in both purified and crude cell extracts with promising results.

Several strategies have been implemented in CFME to increase flux through enzymatic pathways and improve product yield. For example, Opgenorth and coworkers [162] designed a molecular purge valve consisting of an NAD^+-dependent reductase enzyme, an $NADP^+$-dependent reductase enzyme, and a NADH-specific oxidase to manage the flux of reducing equivalents. This valve was utilized to produce valuable compounds such as polyhydroxybutyrate bioplastic and prenylated natural products [162, 163, 228]. The authors carried out efficient synthesis of cannabinoids: up to 1.25 g/L and 1.74 g/L of the cannabinoid precursors, cannabigerolic acid and cannabigerovarinic acid, respectively, were synthesized, then converted in a one-step reaction using cannabidiolic acid synthase, to cannabidiolic acid and cannabidivarinic acid, respectively [228]. The use of co-immobilized enzymes has also been demonstrated in vitro, allowing their reuse over more than seven cycles and highlighting an important step toward large-scale application [30]. In this study, the regulation and identification of inhibitors of the amorpha-4,11-diene (AD) biosynthetic pathway were also carried out, increasing the AD synthesis rate up to 5.7 mol/L·min, an approximately three-fold increase from their previous work (2 mol/L·min) [31]. Other important products such as ethanol [73], n-butanol [113], and ethyl(S)-2-ethoxy-3-(p-methoxyphenyl)propanoate (EEHP) [14], along with next-generation devices such as an aerobic enzymatic fuel cell from glucose have also been demonstrated [262]. A majority of CFME approaches described thus far have used purified enzymes to form biosynthetic pathways. Recently, Dudley and coworkers implemented a novel technique, combining six crude *E. coli* lysates (each enriched with a unique pathway enzyme) at equal ratios and one lysate (enriched with three unique pathway enzymes) to construct a complete 20-step enzymatic pathway of limonene syn-

thesis from glucose [49]. This system achieved a productivity of 3.8 mg/L·h limonene, just two-fold less than the in vivo pathway [9]. In another novel approach, Yi and coworkers [250] used a hybrid system consisting of *E. coli* and cyanobacteria *Synechocystis* sp. PCC6803 cell lysates to demonstrate the synthesis of (R,R)-2,3-butanediol (2,3-BD) from starch, utilizing the presence of starch-degrading enzymes in the cyanobacterial lysate. Thus, cell-free technologies have found wide use for the production of high-value small molecule products, at least in the context of research.

Metabolic engineering approaches have also been used to improve cell-free protein synthesis. Calhoun and Swartz [22], for example, performed chromosomal deletions in the source cells to address the problem of cell-free amino acid degradation. Gene overexpression approaches have also been used to improve protein yields [242, 19]. CFME has also been used to address certain bottlenecks in CFPS such as the need for energy and cofactor regeneration in cell extracts. One of the early examples involved adding oxalate, CoA, and NAD^+ to inhibit a futile cycle while producing ATP from pyruvate [109]. More involved metabolic pathways, including the activation of glycolysis and oxidative phosphorylation, have been utilized in different extracts [98, 22, 23, 96]. These efforts have paved the way for the use of less expensive energy sources (glucose, pyruvate, or glutamate) and nucleotides (NMP), as well as increased the duration of the protein synthesis reactions. Similarly, a novel non-oxidative glycolysis pathway has been designed, enabling 100% conservation of carbon in sugar catabolism to acetyl-CoA [16]. However, despite these achievements and the advantages of cell-free over in vivo processes, a fundamental challenge remains: the optimization of cell-free production systems. Due to the complexity and immense interconnectivity of metabolic networks, even for simple prokaryotic organisms

like *E. coli*, optimizing network operation toward a desired function is often not intuitive [12]. To this end, the systems-level analyses offered by various mathematical modeling tools, developed for application to in-vivo metabolic optimization problems, could prove indispensable for the systematic design of cell-free system operation.

1.4 Mathematical Modeling of Cell-Free Systems

1.4.1 Cell-Free Transcription and Translation Models

If cell-free systems are to become a mainstream technology for advanced applications such as point-of-care therapeutic manufacturing [167], we must first understand the performance limits of these systems [96]. A critical tool towards this goal is mathematical modeling. There have been several mathematical models of cell-free protein synthesis, with the majority of these models exclusively focusing on transcription and translation (TXTL) processes. These models are mostly systems of ordinary differential equations (ODEs) based on saturation or Hill-like kinetic expressions. As an early example, Karzbrun and coworkers developed a coarse-grained model of transcription and translation for *E. coli* cell-free extract [104]. To simplify calculations, this model was based on four enzymes and ten parameters. Transcription and translation processes were assumed to follow Michaelis–Menten kinetics. The authors noted that the protein synthesis rate of their system began to exponentially decay after one hour, so their study focused on the first hour of the cell-free experiment. This decay was attributed to resource depletion and waste accumulation, an important practical consideration regardless of the cell-free sys-

tem used. Stögbauer and coworkers developed a model that accounts for resource consumption and degradation and identified the bottleneck of protein synthesis [207]. Variables representing transcription and translation resources were added to the model, but the exact identities and quantities of these resources were beyond the scope of the study. The authors attempted to use Hill functions to better predict saturation effects of mRNA and their protein of interest, but found that the optimized Hill coefficients were close to one, resulting in Michaelis–Menten-like approximations. Protein yield was determined to be a function of template DNA concentration. This work also found that NTP depletion was not the reason for protein synthesis rate decay; for the specific extract used, ribosome degradation was to blame for rate decay. More recently, Neiß and coworkers published a comprehensive experimentally validated model that identified limiting factors of cell-free protein synthesis [155]. An unusual characteristic of this model was the hybrid black box approach: transcription processes were simplified, while the model for translation was detailed. The entire model was a large system of differential algebraic equations (DAEs); a system of eight algebraic equations and over 400 ODEs. Using sensitivity analysis, Neiß found that cell-free protein synthesis rates were limited by the concentrations of tRNA and elongation factor Tu. A model that captured resource competition in genetic networks was published by Gyorgy and Murry [74]. For a two-protein expression system, simulations that considered both products agreed with the experimental data. This model also predicted possible product concentrations in multiple-protein expression systems and compared different cell-free extracts. The authors concluded that resource competition was a key consideration in the design of synthetic gene circuits. The cell-free protein synthesis models discussed thus far have been

based on experiments in which DNA/protein components were used to construct genetic networks. However, RNA genetic circuitry has also been explored in the cell-free platform, and mathematical models for the system have been developed. Transcriptional regulating RNAs are of interest because they bypass the need for regulatory proteins [123]. In the context of circuit design, regulatory RNAs have been used to create various logic gates and cascades [20, 29]. The first experimentally validated model of a synthetic RNA circuit was published by Hu and coworkers [86]. The model contained eight ODEs and 13 previously unknown parameters. These parameters were estimated based on results from sensitivity analysis-guided experimental design. Taken together, models of transcription, translation, resource competition, and gene regulatory circuits have provided useful information for optimizing cell-free biomanufacturing or designing new systems; however, they have each provided an incomplete representation of cell-free protein synthesis. CFPS does not just rely on transcription and translation processes, but instead depends on central carbon metabolism and other metabolic pathways to meet energy and carbon precursor requirements. Thus, more sophisticated models are needed that integrate metabolic pathways with transcription and translation processes. Ultimately, an integrated framework could provide insights into the limitations of CFPS and generate strategies for improving performance metrics such as carbon yield, energy efficiency, and productivity.

1.4.2 Metabolic Modeling Frameworks

Traditional approaches to metabolic modeling, which were first developed to describe living cells, could also be applied to cell-free systems, thereby ad-

dressing an important current limitation. Decades before the genomics revolution, mechanistically structured in vivo metabolic models arose from the desire to predict microbial phenotypes resulting from changes in intracellular or extracellular states [54]. The single-cell *E. coli* models of Shuler and coworkers pioneered the construction of large-scale, dynamic metabolic models that incorporated multiple regulated catabolic and anabolic pathways constrained by experimentally-determined kinetic parameters [45]. Shuler and coworkers generated many single-cell kinetic models, including single-cell models of eukaryotes [205, 247], minimal cell architectures [26], and DNA sequence-based whole-cell models of *E. coli* [10]. As biological understanding grew, metabolic models became less reductionist and more detailed. Next-generation models described cellular processes such as RNA synthesis, chromosome synthesis, and regulated catabolic and macromolecular synthesis pathways in detail using ordinary differential equations [224]. For example, Karr et al. (2012) developed a whole cell model of *Mycoplasma genitalium*, accounting for all genes and their interactions in the cell [102]. The model, which was constructed with independent sub-models describing different components of the cell, successfully described the full cellular life cycle at the level of single molecules. However, while undoubtedly important tools, traditional metabolic modeling approaches are often complex and nonlinear and require the estimation of a large number of unknown parameters; a difficult process because of the inherent noisiness of biological data and the computational burden of repeatedly solving the model equations. To overcome such obstacles, constraint-based methods were developed to describe metabolic networks with only a limited need for kinetic parameters [229].

Stoichiometric reconstructions of microbial metabolism, popularized

by constraint-based approaches such as flux balance analysis (FBA), have become standard tools to interrogate metabolism [117]. FBA and metabolic flux analysis (MFA) [243], as well as convex network decomposition approaches such as elementary modes [188] and extreme pathways [187] model intracellular metabolism using the biochemical stoichiometry and other constraints such as thermodynamic feasibility [80, 77] under pseudo-steady-state conditions. Constraint-based approaches use linear programming [35] to predict productivity [229, 185], yield [229], mutant behavior [51], and growth phenotypes [160] for biochemical networks of varying complexity, including genome-scale networks. Constraint-based models have also been used to identify strategies for the overproduction of desired compounds. These strategies include genetic knockouts or the addition of heterologous enzyme pathways to an organism's metabolic network and have been used in developing useful bacterial strains for the production of biofuels [11], high-value chemicals [150, 202, 251], and pharmaceuticals [180, 53]. Stoichiometric reconstructions have been expanded to include the metabolic demands for protein synthesis. Allen and Palsson developed sequence-specific constraint-based models, based on the DNA and protein sequences of interest, where transcription and translation processes were integrated with metabolism [8]. Since the early work of Allen and Palsson, sequence-specific constraint-based models have been expanded to the genome scale with detailed descriptions of gene expression (ME-model) [221, 117, 160] and protein structures (GEM-PRO) [259, 28]. These expansions have greatly increased the scope of questions stoichiometric reconstructions can explore. For example, constraint-based methods, which are powerful tools to estimate the performance of metabolic networks, could potentially predict nonintuitive strategies to optimize the interaction between metabolism and gene expression

in cell-free applications. Thus, the use of integrated constraint-based models for cell-free optimization studies is a promising future research direction.

1.4.3 Emergence of Integrated Cell-Free Models

Modeling the integration of cell-free transcription and translation processes with metabolic pathways remains in its infancy, with only a few published mathematical models [39, 236, 85]. Horvath and coworkers developed an ensemble of dynamic *E. coli* CFPS models using parameters estimated from measurements of metabolite, amino acid, and protein concentrations from CFPS reactions conducted using the PANOx-SP system [85]. This work built upon the hybrid cell-free modeling approach of Wayman and colleagues, which integrated kinetic modeling with a rule-based description of allosteric control [241]. By simulating reaction group knockouts, Horvath et al. suggested that cell-free metabolism and protein synthesis were strongly coupled with oxidative phosphorylation and glycolytic flux. On the other hand, to circumvent computationally expensive parameter estimation, Vilkhovoy and coworkers [236] developed an experimentally-validated constraint-based model of CFPS, which integrated the expression of a model protein product with the supply of metabolic precursors and energy (Figure 1.3). This model coupled transcription and translation processes with available resources using only six adjustable parameters. Model analysis suggested that protein expression in the PANOx-SP system was translationally limited. Further, the same modeling approach, using only a limited number of experimentally-derived parameters, also described protein expression in the myTXTL system, thereby underscoring the power and versatility of the approach. Taken together, the incorporation of complex metabolism with

genetic regulatory networks using constraint-based modeling is a promising approach to simulate cell-free systems. Unfortunately, despite these early studies, there remains an unmet need for comprehensive metabolic models of cell-free reactions. However, as experimental methods are developed for cell-free systems, e.g., [235], and benchmark cell-free datasets are published, we expect that the metabolic modeling community will address this shortcoming.

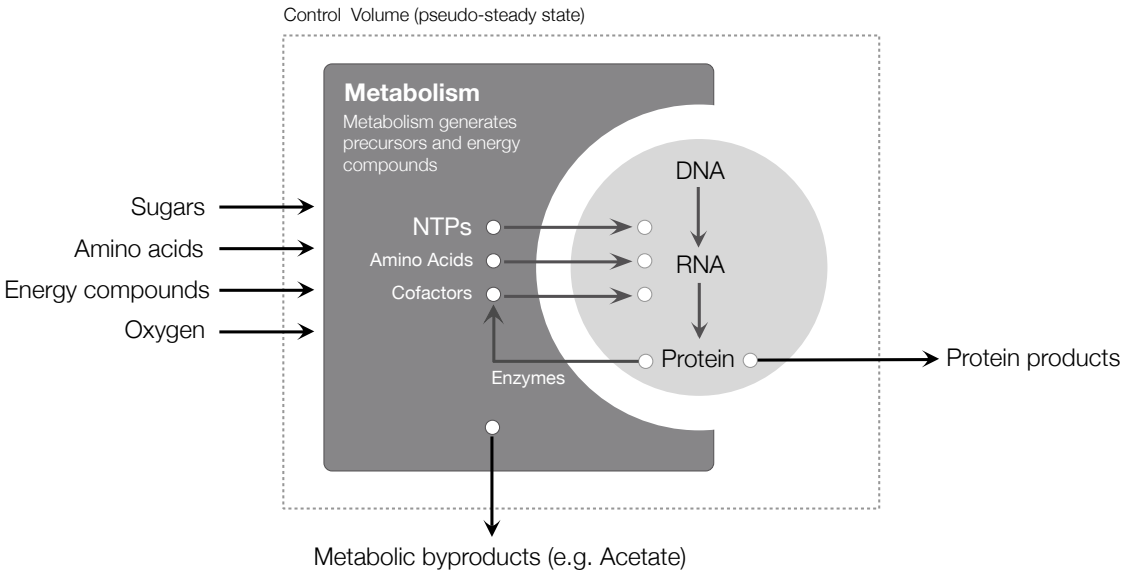


Figure 1.3: Schematic of the integration of transcription and translation processes with metabolism. Transcription and translation processes demand macromolecular precursors (e.g., NTPs, amino acids, and cofactors) from metabolism for gene expression. The integrated framework is represented as a stoichiometric matrix of metabolites participating in certain reactions, along with a description of the metabolic demands for protein expression. The metabolic flux is estimated subject to constraints, a pseudo-steady-state assumption, and an objective function.

1.5 Conclusions

Cell-free systems have evolved from an investigative tool used since the early 1950s to a sophisticated platform useful for a variety of bioengineering, bioman-

ufacturing, and synthetic biology applications. With the recent advances in extract preparation, improvements in energy regeneration mechanisms, and the ability to perform high-throughput continuous reactions in microfluidic chips, cell-free systems are emerging as a viable alternative to traditional living cells in several application areas. For example, they have become valuable investigative tools for metabolic engineering research, given their tunability and the unfettered access to metabolism without the interference of the cell wall. However, the optimization of cell-free applications remains an important challenge. Toward this challenge, mathematical and computational modeling is a critical tool that could move the platform forward. Models facilitate the understanding of the role of systems-level parameters and interactions, and they also promote the generation of metabolic engineering strategies, for example assisting in making the appropriate genetic manipulation for a desired function, which is not always intuitive. The broad review of the literature presented here highlighted several experimental and computational opportunities that could be addressed in future work. For example, a more detailed description of transcription and translation reactions has been utilized in genome-scale metabolic engineering models, e.g., O'Brien et al. [160]. These template reactions could be adapted to a cell-free system, allowing us to consider important facets of protein production, such as the role of chaperones in protein folding. In addition, post-translational modifications such as glycosylation that are important for the production of therapeutic proteins could also be included in the next generation of cell-free models. Finally, constraint-based modeling could be extended to multi-protein synthetic circuits, RNA circuits, or small molecule production. There are also opportunities to explore with regard to the preparation and manipulation of cell extracts. For example, enabling selec-

tive enzyme deletions directly in cell-free extracts, without having these deletions stem from the extract preparation process, could be a game-changing technology. Taken together, continual advancements in modeling and experimental design have paved the way for cell-free systems to become valuable tools for molecular biology research and a promising platform for manufacturing of valuable biotherapeutics and chemicals.

However, while the popularity of cell-free systems has grown dramatically in the research community, the platform still faces important challenges for biomanufacturing applications, in particular scale-up, the high cost of extracts, and limited post-translational modification capability. While mammalian and insect cell-free systems have post-translational modification machinery, these types of extracts are expensive compared to their bacterial counterparts. Toward this challenge, DeLisa and coworkers [71, 95] recently expressed N-linked glycoproteins in *E. coli* cell-free extracts, opening up possibilities for the production of therapeutically-relevant proteins in bacterial extracts, which have better overall protein yields, are relatively inexpensive, and have easier extract preparation protocols. There have also been studies on the possible scale-up of protein synthesis by several research groups and a few startup biotechnology companies. However, this remains a critical and underserved topic area. Another possible challenge to using cell-free systems for applications such as biosensors, on-demand therapeutics, or even industrial production, is the inability to achieve 100% lysis of cells during the extract preparation process, resulting in residual contaminating cells. Addressing this concern could be important to meet various Food and Drug Administration (FDA) regulations. Toward this challenge, Smith and coworkers assessed decontamination strategies and showed that sterile filtration and lyophilization effectively removes cell contam-

ination without affecting the protein synthesis capabilities of the system [198]. Thus, while there remain important challenges to be overcome, it is encouraging that several research groups have worked toward addressing the challenges faced by the the platform. However, continued research and development is important to further improve the capabilities of cell-free technology.

1.6 Acknowledgments

The work described was supported by the Center on the Physics of Cancer Metabolism through Award Number 1U54CA210184-01 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

1.7 Abbreviations

The following abbreviations are used in this manuscript:

CFPS	cell-free protein synthesis
TXTL	transcription and translation
DNA	deoxyribonucleic acid
RNA	ribonucleic acid
NMP	nucleoside monophosphate
NTP	nucleoside triphosphate
ATP	adenosine triphosphate
GTP	guanosine triphosphate
NAD	nicotinamide adenine dinucleotide
CoA	coenzyme A
mRNA	messenger RNA
tRNA	transfer RNA
PURE	Protein synthesis Using Recombinant Elements
ODE	ordinary differential equation
CFME	cell-free metabolic engineering
FBA	flux balance analysis
MFA	metabolic flux analysis
ME	metabolic engineering
PCR	polymerase chain reaction

CHAPTER 2

ABSOLUTE QUANTIFICATION OF CELL-FREE PROTEIN SYNTHESIS METABOLISM BY REVERSED-PHASE LIQUID CHROMATOGRAPHY-MASS SPECTROMETRY

2.1 Abstract

[1] Cell-free protein synthesis (CFPS) is a widely used research tool in systems and synthetic biology; however, if CFPS is to become a mainstream technology for applications such as point-of-care manufacturing, we must understand the performance limits of these systems. Toward this question, we developed a robust protocol to quantify 40 compounds involved in glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, energy metabolism and cofactor regeneration in CFPS reactions. The method uses internal standards tagged with 13C-aniline, while compounds in the sample are derivatized with 12C-aniline. The internal standards and sample were mixed and analyzed by reversed-phase liquid chromatography-mass spectrometry (LC/MS). The co-elution of compounds eliminated ion suppression, allowing the accurate quantification of metabolite concentrations over 2-3 orders of magnitude where the average correlation coefficient was 0.988. Five of the forty compounds were untagged with aniline, however they were still detected in the CFPS sample and quantified with a standard curve method. The chromatic run takes approximately 10 minutes to complete. In summary, we developed a fast, robust method to separate, and accurately quantify 40 compounds involved in CFPS

[1] Adapted with permission from Vilkhovoy, M.; Dai, D.; Vadhin, S.; Adhikari, A.; Varner, J. D. Absolute Quantification of Cell-Free Protein Synthesis Metabolism by Reversed-Phase Liquid Chromatography-Mass Spectrometry. JoVE 2019, No. 152, 60329.

in a single LC/MS run. Taken together, the method is a robust and accurate approach to characterize cell free metabolism, so that ultimately, we can understand and improve the yield, productivity and energy efficiency of cell free systems.

2.2 Introduction

Cell-free protein synthesis has become a widely used tool in systems and synthetic biology, and a promising technology for point-of-use manufacturing of biomolecules. Cell-free systems offer many advantages compared to in vivo processes, such as direct access to metabolites and the biosynthetic machinery without the interference of a cell wall or the complications associated with cell growth [83]. However, a fundamental understanding of the performance limits of cell free processes has been lacking. High-throughput methods for metabolite quantification are valuable because they can help characterize metabolism, they are important to our understanding of the systems, and are critical to the construction of robust metabolic computational models useful in process optimization[236, 232, 85]. Common methods used to determine metabolite concentrations include Nuclear Magnetic Resonance (NMR), Fourier transform-infrared spectroscopy (FT-IR), enzyme-based assays, and mass spectrometry (MS)[75, 41, 182, 220]. However, these methods are often limited by their inability to efficiently measure multiple compounds at once and sample size requirements. For example, enzyme-based assays can often only be used to quantify a single compound in a run, and are limited when the sample size is small, such as in cell-free protein synthesis reactions (typically run on a 10-15 μL scale). Meanwhile, NMR requires a high abundance of metabolites for

detection and quantification[41]. Toward these shortcomings, chromatography methods in tandem with mass spectrometry (LC/MS) provide several advantages, including sensitivity and the capability of measuring multiple species simultaneously[50]; however, the analytical complexity increases considerably with the number and diversity of species being measured. It is important, therefore, to develop methods that fully realize the high-throughput potential of LC/MS systems. Compounds in a sample are separated by liquid chromatography and identified through mass spectrometry. The signal of the compound depends on its concentration and ionization efficiency, where the ionization can vary between compounds and may also depend on the sample matrix.

Achieving the same ionization efficiency between the sample and standards is a challenge to using LC/MS to quantify analytes. Further, quantification becomes more challenging with metabolite diversity due to signal splitting and heterogeneity in proton affinity and polarity[88]. Lastly, the co-eluting matrix of the sample can also affect the ionization efficiencies of the compounds. To address these issues, metabolites can be chemically derivatized, increasing the separation resolution, and the sensitivity and detection by the LC/MS system, while simultaneously decreasing signal splitting in some cases[88, 87]. Chemical derivatization works by tagging specific functional groups of metabolites to adjust their physical properties like charge or hydrophobicity to increase ionization efficiency[87]. Various tagging agents can be used to target different functional groups like amines, hydroxyls, phosphates, carboxylic acids, etc. Aniline, one such derivatization agent, targets multiple functional groups at once, and adds a hydrophobic component into hydrophilic molecules, increasing their separation resolution and signal[248]. To address the co-eluting matrix ion suppression effect, Yang and coworkers developed a technique based on

Group Specific Internal Standard Technology (GSIST) labeling where standards are tagged with 13C aniline isotopes and mixed with the sample[248, 94]. The metabolite and corresponding internal standard have the same ionization efficiency since they co-elute, and their intensity ratio can be used to quantify the concentration in the experimental sample.

In this study, we developed a protocol to detect and quantify 40 compounds involved in glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, energy metabolism and cofactor regeneration in cell-free protein synthesis reactions. The method is based on the GSIST approach, where we used 12C-aniline and 13C-aniline to tag, detect, and quantify metabolites using reversed-phase LC/MS. The linear range of all compounds spanned 2-3 orders of magnitude with an average correlation coefficient of 0.988. Thus, the method is a robust and accurate approach to interrogate cell free metabolism, and possibly whole-cell extracts.

2.3 Representative Results

As a proof-of-concept, we used the protocol to quantify metabolites in myTXTL, a commercially available E. coli based CFPS system (Arbor Biosciences) expressing green fluorescent protein (GFP). The CFPS reaction (14μL) was quenched and de-proteinized with ethanol. The CFPS sample was then tagged with 12C-aniline, while standards were tagged with 13C-aniline. The tagged sample and standards were then combined and injected into the LC/MS (Fig. 2.1). The protocol detected and quantified 40 metabolites involved in central carbon and energy metabolism using internal standards, while a standard curve for 5 of the

metabolites that were not tagged with aniline was also developed (Fig. 2.2 and Table 2.1). The diverse metabolites involved in these pathways were a class of phosphorylated sugars, phosphocarboxylic acids, carboxylic acids, nucleotides, and cofactors. The derivatization with aniline introduced a hydrophobic moiety into hydrophilic molecules which facilitated more effective separation using reversed-phase chromatography[248]. In addition, the method enabled the separation of structural isomer pairs such as glucose 6-phosphate and fructose 6-phosphate in a single LC/MS run. Each compound's mass over charge (m/z) ratio and retention time were identified prior to the experiment by injecting 1mM of one compound at a time and comparing the mass spectrum to the blank (Table 2.2).

The limit of detection and range of linearity for all compounds was estimated by producing a standard curve that ranged from 0.10 μM to 400 μM (Table 2.1). The average correlation coefficient (R^2) for all compounds was 0.988 and most compounds had a linear range of 3-orders of magnitude. Three compounds had notable saturation effects, especially alpha-ketoglutarate which had a linear range from 0.1 μM to 25 μM. Isocitrate and citrate also had saturation effects above 100 μM.

2.4 Materials and Methods

Preparation of reagents for aniline tagging

A 6M aniline solution at pH 4.5 was prepared as follows: 550 μL of aniline was mixed with 337.5 μL of UltraPure water and 112.5 μL of 12 M HCl (hydrochlo-

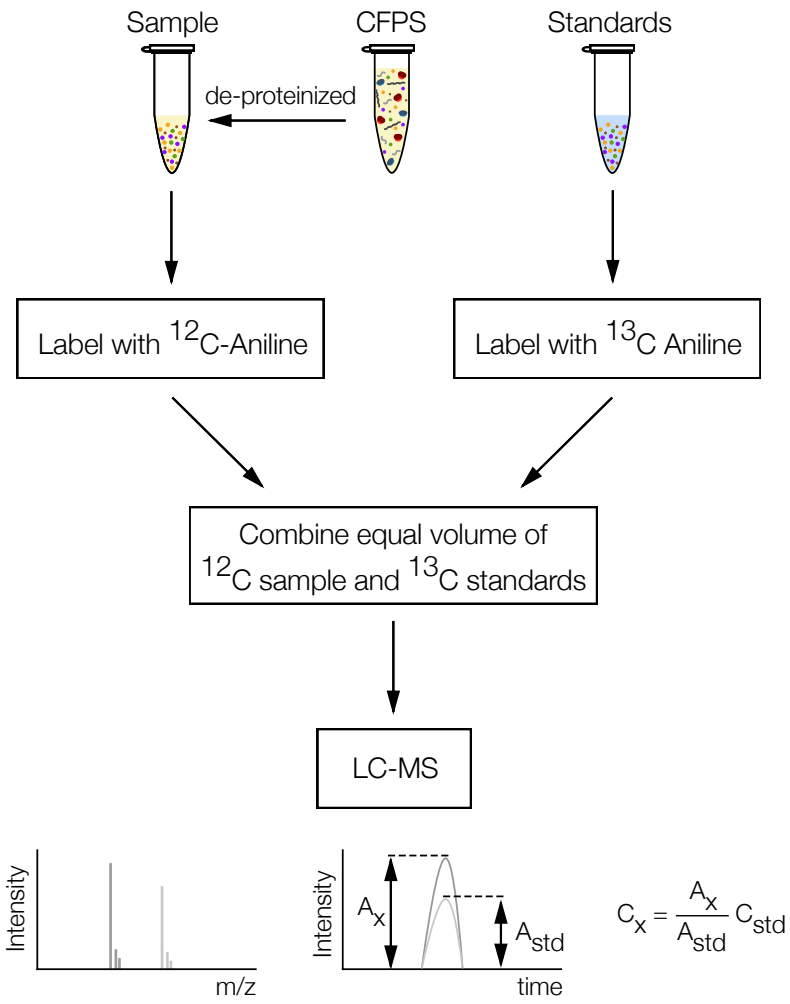
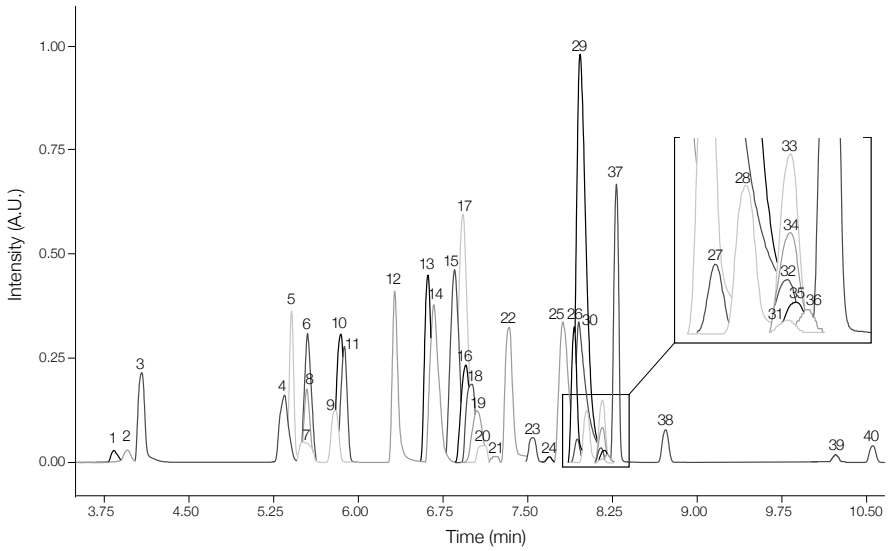


Figure 2.1: Schematic of workflow for aniline tagging. The cell-free protein synthesis reaction is de-proteinized and tagged with 12C-aniline, while a standard stock mixture is tagged with 13C-aniline. Both mixtures are then mixed at a 1:1 volumetric ratio and analyzed by LC/MS.

ric acid) in a centrifuge tube. The solution was vortex well and stored at 4 °C. 6M 13C aniline solution was prepared at pH 4.5 by combining 250 mg of 13C-aniline with 132 μL of water and 44 µL of 12 M HCl. The solution was vortexed well and stored at 4 °C. 200 mg/mL EDC (N-(3-dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride) solution was prepared by dissolving 2 mg EDC in 10 μL water for every sample to be tagged.



1. Gly3P	9. LAC	17. UDP	25. CTP	33. MAL
2. NAD	10. AMP	18. FAD	26. GTP	34. GAP
3. GLC	11. UMP	19. F16P	27. OAA	35. ACA
4. S7P	12. NADP	20. 6PG	28. aKG	36. NADPH
5. F6P	13. 3PG	21. NADH	29. UTP	37. PEP
6. GMP	14. CDP	22. G6P	30. ATP	38. SUCC
7. RL5P	15. GDP	23. R5P	31. FUM	39. ICIT
8. CMP	16. ADP	24. E4P	32. PYR	40. CIT

Figure 2.2: Overlapped selected ion chromatograms for 40 metabolites. Mass chromatogram from a single LC/MS run of a 40μM standard mixture of 40 metabolites. Peaks were identified by their retention time and m/z values for each compound. Complete compound names and their abbreviations are listed in Table 2.1.

Preparation of standards

All compounds except for NAD, NADP, FAD, acetyl CoA, and glycerol 3-phosphate, were mixed to create a 2 mM stock solution of all compounds. Separately, NAD, NADP, FAD, acetyl CoA, and glycerol 3-phosphate with the appropriate volumes were mixed to create a 2 mM stock solution.

Preparation of Sample

The proteins in a cell-free protein synthesis reaction were quenched and precipitated by adding an equal volume of ice-cold 100

Labeling reaction

The samples were labelled with 12C as follows. 6 μL of sample was transferred into a new centrifuge tube and the volume was brought to 50 μL with water. 5 μL of 200mg/ml EDC and 12C solution were added, and the reaction vortexed with gentle shaking for 2 hours at room temperature. After 2 hours, the tubes were removed from the shaker and 1.5 μL of TEA (triethylamine) was added to the reaction in a fume hood. The mixture was centrifuged at 13,500 x g for 3 minutes. The 25 μL each of the tagged internal standard and tagged samples were mixed, and then analyzed by the LCMS system. In order to create a standard curve for the untagged metabolites, the stock solution was tagged following a similar procedure, and then analyzed by the LCMS system.

Quantification of metabolites

Concentration of the unknown compound was calculated using:

$$C_{x,i} = \frac{A_{x,i}}{A_{std,i}} C_{std,i} D \tag{2.1}$$

where $C_{x,i}$ is the concentration of the unknown sample for metabolite i, $A_{x,i}$ is the integrated area of the unknown metabolite i, $A_{std,i}$ is the integrated area of the internal standard of metabolite i, $C_{std,i}$ is the concentration of the internal standard of metabolite i, and D is the dilution factor.

Untagged metabolites were quantified using the standard curve method.

2.5 Acknowledgements

The work described was supported by the Center on the Physics of Cancer Metabolism through Award Number 1U54CA210184-01 from the National Cancer Institute (https://www.cancer.gov/). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Table 2.1: Each compound's corresponding limit of detection, range of linearity and correlation coefficient identified from standard curves.

Peak	Metabolite	Abbreviation	KEGG ID	Limit of Detection (μM)	Limit of Linear Range (μM)	R^2
1	Glycerol 3-phosphate	Gly3P	C00093	0.1	400	0.995
2	Nicotinamide adenine dinucleotide	NAD	C00003	0.39	400	0.993
3	Glucose	GLC	C00031	0.1	400	0.997
4	Sedoheptulose 7-phosphate	S7P	C05382	0.16	400	0.988
5	Fructose 6-phosphate	F6P	C00085	0.1	400	0.986
6	Guanosine monophosphate	GMP	C00144	0.39	100	0.992
7	Ribulose 5-phosphate	RL5P	C00199	0.39	400	0.996
8	Cytidine monophosphate	CMP	C00055	0.1	100	0.992
9	Lactate	LAC	C00186	0.1	400	0.988
10	Adenosine monophosphate	AMP	C00020	0.1	100	0.992
11	Uridine monophosphate	UMP	C00105	0.1	100	0.997
12	Nicotinamide adenine dinucleotide phosphate	NADP	C00006	0.34	400	0.950
13	3-Phosphoglyceric acid	3PG	C00197	0.1	100	0.996
14	Cytidine diphosphate	CDP	C00112	0.39	400	0.997
15	Guanosine diphosphate	GDP	C00035	1.5625	400	0.984
16	Adenosine diphosphate	ADP	C00008	0.39	400	0.995
17	Uridine diphosphate	UDP	C00015	0.39	400	0.991
18	Flavin adenine dinucleotide	FAD	C00016	0.1	400	0.958
19	Fructose 1,6-bisphosphate	F16P	C05378	0.39	400	0.989
20	Gluconate 6-phosphate	6PG	C00345	0.39	400	0.989
21	Nicotinamide adenine dinucleotide reduced	NADH	C00004	0.39	100	0.972
22	Glucose 6-phosphate	G6P	C00668	0.1	400	0.984
23	Ribose 5-phosphate	R5P	C00117	0.39	100	0.999
24	Erythrose 4-phosphate	E4P	C00279	0.39	400	0.979
25	Cytidine triphosphate	CTP	C00075	6.25	100	0.998
26	Guanosine triphosphate	GTP	C00044	6.25	100	0.993
27	Oxalacetate	OAA	C00036	0.56	400	0.997
28	Alpha-ketoglutarate	aKG	C00026	0.1	25	0.979
29	Uridine triphosphate	UTP	C00075	1.5625	400	0.998
30	Adenosine triphosphate	ATP	C00002	1.5625	400	0.991
31	Fumarate	FUM	C00122	1.5625	100	0.999
32	Pyruvate	PYR	C00022	0.39	400	0.993
33	Malate	MAL	C00149	0.1	400	0.991
34	D-glyceraldehyde 3-phosphate	GAP	C00118	0.1	100	0.974
35	Acetyl-coenzyme A	ACA	C00024	0.1	100	0.991
36	Nicotinamide adenine dinucleotide phosphate reduced	NADPH	C00005	0.14	100	0.990
37	Phosphoenolpyruvate	PEP	C00074	0.1	100	0.962
38	Succinate	SUCC	C00042	0.1	320	0.999
39	Isocitrate	ICIT	C00311	0.39	100	0.998
40	Citrate	CIT	C00158	0.1	100	0.981

Table 2.2: Each compound's corresponding peak number, retention time, m/z value for 12C, 13C, and unlabeled, cone voltage, and MS species.

Peak	Metabolite	KEGG ID	Retention Time (min)	12C m/z	13C m/z	nonlabel m/z	CV	MS Species
1	Gly3P	C00093	3.85			153	10	M – H2O – H
2	NAD	C00003	3.96			698	10	M + Cl – H
3	GLC	C00031	4.06	289.9	296		15	M + A + Cl - H
4	S7P	C05382	5.41	364	370		10	M + A – H
5	F6P	C00085	5.48	334	340		10	M + A – H
6	GMP	C00144	5.57	437.05	443		10	M + A – H
7	RL5P	C00199	5.58	304	310		10	M + A – H
8	CMP	C00055	5.59	397.09	403		10	M + A – H
9	LAC	C00186	5.77	164.05	170		10	M + A – H
10	AMP	C00020	5.85	421.1	427.1		10	M + A – H
11	UMP	C00105	5.88	398.07	404		10	M + A – H
12	NADP	C00006	6.39			724	10	M - H2O – H
13	3PG	C00197	6.63	242	248.06		15	M + A – H2O – H
14	CDP	C00112	6.72	477	483		10	M + A – H
15	GDP	C00035	6.87	517	523		10	M + A – H
16	ADP	C00008	6.94	501	507		10	M + A – H
17	UDP	C00015	6.97	478	484		10	M + A – H
18	FAD	C00016	7.03			784.15	15	M – H
19	F16P	C05378	7.1	395.95	402.1		10	M + A – H2O – H
20	6PG	C00345	7.11	425.1	437		10	M + 2A – H
21	NADH	C00004	7.23	633.13	639.08		10	M + A + H2O – nicotinamide – H
22	G6P	C00668	7.32	409.1	421.1		10	M + 2A – H
23	R5P	C00117	7.54	379.1	391.1		15	M + 2A – H
24	E4P	C00279	7.71	348.9	361		10	M + 2A – H
25	CTP	C00075	7.84	557	563		5	M + A – H
26	GTP	C00044	7.93	597	603		5	M + A – H
27	OAA	C00036	7.94	281	293		25	M + 2A – H
28	aKG	C00026	7.95	295	307.1		15	M + 2A – H
29	UTP	C00075	7.97	558	564		10	M + A – H
30	ATP	C00002	8.03	581	587		15	M + A – H
31	FUM	C00122	8.09	265	277.1		10	M + 2A – H
32	PYR	C00022	8.09	162	168		25	M + A – H
33	MAL	C00149	8.09	283.06	295.15		10	M + 2A – H
34	GAP	C00118	8.09	319	331.1		5	M + 2A – H
35	ACA	C00024	8.16			790	10	M – H2O – H
36	NADPH	C00005	8.23	694.92	700.82		10	M + A – nicotinamide – H
37	PEP	C00074	8.28	317	329.1		20	M + 2A – H
38	SUCC	C00042	8.64	267.07	279.1		15	M + 2A – H
39	ICIT	C00311	10.13	398	416		10	M + 3A – H2O – H
40	CIT	C00158	10.46	416.1	434.06		20	M + 3A – H

CHAPTER 3

EFFECTIVE BIOPHYSICAL MODELING OF CELL FREE TRANSCRIPTION AND TRANSLATION PROCESSES

3.1 Abstract

[1] Transcription and translation are at the heart of metabolism and signal transduction. In this study, we developed an effective biophysical modeling approach to simulate transcription and translation processes. The model, composed of coupled ordinary differential equations, was tested by comparing simulations of two cell free synthetic circuits with experimental measurements generated in this study. First, we considered a simple circuit in which sigma factor 70 induced the expression of green fluorescent protein. This relatively simple case was then followed by a more complex negative feedback circuit in which two control genes were coupled to the expression of a third reporter gene, green fluorescent protein. Many of the model parameters were estimated from previous biophysical studies in the literature, while the remaining unknown model parameters for each circuit were estimated by minimizing the difference between model simulations and messenger RNA (mRNA) and protein measurements generated in this study. In particular, either parameter estimates from published studies were used directly, or characteristic values found in the literature were used to establish feasible ranges for the parameter estimation problem. In order to perform a detailed analysis of the influence of individual model parameters on the expression dynamics of each circuit, global sensitivity analy-

[1] Adapted with permission from Adhikari, A.; Vilkhovoy, M.; Vadhin, S.; Lim, H. E; Varner, J. D. Effective Biophysical Modeling of Cell Free Transcription and Translation Processes. Frontiers in Bioengineering and Biotechnology. 8, 539081 (2020).

sis used. Taken together, the effective biophysical modeling approach captured the expression dynamics, including the transcription dynamics, for the two synthetic cell free circuits. While we considered only two circuits here, this approach could potentially be extended to simulate other genetic circuits in both cell free and whole cell biomolecular applications as the equations governing the regulatory control functions are modular and easily modifiable. The model code, parameters, and analysis scripts are available for download under an MIT software license from the Varnerlab GitHub repository.

3.2 Introduction

Cell free systems are a widely used research tool in systems and synthetic biology and a promising platform for the manufacturing of proteins and chemicals [234]. A distinctive feature of cell free systems is the absence of cellular growth and maintenance, thereby allowing the direct allocation of carbon and energy resources toward a product of interest. Cell free systems are also more amenable than living systems to observation and manipulation, hence allowing rapid tuning of reaction conditions. Arguably, the most widely used cell free technology is cell free protein synthesis (CFPS), an *in vitro* platform for protein transcription (TX) and translation (TL). Transcription and translation, the processes by which information stored in DNA is converted to a working protein, are at the center of metabolism and signal transduction programs important to biotechnology and human health. For example, evolutionarily conserved developmental programs such as the epithelial to mesenchymal transition (EMT) [222], or retinoic acid induced differentiation [156], rely on multiple rounds of highly coordinated gene expression. From the perspective of biotechnology, even relatively simple

industrially important organisms such as *Escherichia coli*, have intricate transcriptional regulatory networks which control the metabolic state of the cell in response to changing nutrient conditions [164, 233]. Understanding the dynamics of regulatory networks can be greatly facilitated by mathematical models. A majority of these models fall into three categories: logical, continuous, and stochastic models [101]. Logical models such as Boolean networks [65] developed using a variety of approaches and data [173] represent the state of each network entity as a discrete variable, provide a quick but qualitative description of the behavior of the regulatory network. Linear and non-linear ordinary differential equation (ODE) models fall into the second category, and they generally provide a detailed picture of the network dynamics, although they can be non-physical models, e.g., relying on a gene signal perspective [17]. Lastly, stochastic models describe the interactions between individual molecules, and discrete reaction events [135, 138, 107, 177]. Model choice depends on criteria such as speed, the level of detail required and the quantity of experimental data available to estimate the model parameters. While the end goal of the models might be to accurately predict *in vivo* behavior, living systems do not necessarily provide an ideal experimental platform. For example, although there have been significant advancements in metabolomics e.g., [168], the rigorous quantification of intracellular messenger RNA (mRNA) copy number or protein abundance remains challenging. Toward this challenge, cell free systems offer several advantages for the study of transcription and translation processes.

Cell free biology has historically been an important tool to study the fundamental biological mechanisms involved with gene expression. In the 1950s, cell free systems were used to explore the incorporation of amino acids into proteins [18, 245, 246], and the role of adenosine triphosphate (ATP) in protein

production [82]. Further, *E. coli* extracts were used by Nirenberg and Matthaei in 1961 to demonstrate templated translation [157, 132], leading to a Nobel Prize in 1968 for deciphering the codon code. More recently, as advancements in extract preparation and energy regeneration have extended their durability, the usage of cell free systems has also expanded to both small- and large-scale biotechnology and biomanufacturing applications [211, 196]. Today, cell free systems have been implemented for therapeutic protein and vaccine production [95, 204, 151], biosensing [200], genetic part prototyping [144] and minimal cell systems [257]. The versatility of cell free systems offers a tremendous opportunity for the systems-level experimental and computational study of biological mechanism. For example, a number of ordinary differential equation based cell free models have been developed [208, 134, 131, 44, 128]. However, despite the obvious advantages offered by a cell free system, experimental determination of the kinetic parameters for these models is often challenging. For instance, the cell free modeling study of Horvath and coworkers (which included a description of transcription and translation, and the underlying metabolism supplying energy and precursors for transcription and translation), had over 800 unknown model parameters [84]. Moreover, the construction, identification and validation of the Horvath model took well over a year to complete. Thus, constructing, identifying and validating biophysically motivated cell free models, which are simultaneously manageable, is a key challenge. Toward this challenge, effective modeling approaches which coarse grain biological details but remain firmly rooted in a biophysical perspective, could be an important tool.

In this study, we developed an effective biophysical modeling approach to simulate cell free transcription and translation processes. The model used classical biophysical arguments to formulate kinetic expressions for the rates of

transcription and translation. These rates were then used in material balance equations (ordinary differential equations) to simulate the mRNA and protein concentration as a function of time for different cell free genetic circuits. The model was effective as it neglected potentially important mechanistic factors, and the integration of transcription and translation with metabolism. For example, the model did not consider how the transcription and translation rate was influenced by the availability of metabolic resources, e.g., energy or building block concentration. Nor did the model consider potentially important biology, for example the role of elongation factors or protein folding chaperones (among many other potentially important factors). We tested this approach by comparing simulations of two cell free synthetic circuits with messenger RNA (mRNA) and protein measurements (deGFP) generated in this study using the *E. coli* based myTXTL cell free system. First, we considered a simple circuit ($C1$) in which endogenous sigma factor 70 (σ_{70}) induced the expression of a fast maturing dual emission green fluorescent protein variant (deGFP). This relatively simple case was then followed by a more complex negative feedback circuit ($C2$) where two control genes were coupled to the expression of deGFP. The second circuit is an extension of the first, with the presence of additional regulatory elements. Characteristic values for many of the model parameters were estimated from published biophysical studies or took the form of corrections to order of magnitude literature estimates, while the remaining unknown model parameters for each circuit were estimated by minimizing the difference between simulated and measured mRNA and protein concentrations. In particular, either parameter estimates from published studies were used directly, or characteristic values found in the literature were used to establish feasible ranges for the parameter estimation problem. Next, in order to provide a detailed insight into

the influence of individual model parameters on the expression dynamics of each circuit, Morris sensitivity analysis was employed. For $C1$, the sensitivity results were informative, but expected. However, for $C2$, the analysis hierarchically stratified the parameters (and associated model species) into local versus global categories. For example, parameters that controlled the abundance of lambda phage repressor protein (cI-ssrA), a master circuit regulator in $C2$, were globally important as they influenced all other species. On the other hand, the parameters that influenced deGFP levels (the endpoint of both circuits) were only locally important to deGFP. Taken together, the effective biophysical modeling approach captured the expression dynamics, including the transcription dynamics, for two synthetic cell free circuits. While we considered only two circuits here, this approach could potentially be extended to simulate other genetic circuits in both cell free and whole cell biomolecular applications. The model code, parameters, and analysis scripts are available under an MIT software license from the Varnerlab GitHub repository [231].

3.3 Materials and Methods

3.3.1 Cell free protein synthesis reactions.

The cell free protein synthesis (CFPS) reactions were carried out using the myTXTL Sigma 70 Master Mix (Arbor Biosciences) in 1.5 mL Eppendorf tubes. The working volume of all the reactions was 12 μL, composed of the Sigma 70 Master Mix (9 μL) and the plasmids (3 μL total): P70a-deGFP (5 nM) for the single-gene system; P70a-deGFP-ssrA (8 nM), P70a-S28 (1.5 nM) and

P28a-cI-ssrA (1 nM) for the negative feedback circuit. The plasmids were bought in lyophilized form (Arbor Biosciences) and purified using QIAprep Spin Miniprep Kit (Qiagen) using cell lines DH5-Alpha (for P28a-cI-ssrA) or KL740 (for P70a-deGFP, P70a-deGFP-ssrA, and P70a-S28). The CFPS reactions were incubated at 29°C.

3.3.2 mRNA and protein quantification.

Following each CFPS run, the total RNA was extracted from 1 μL of the reaction mixture using PureLink RNA Mini Kit (Thermo Fisher Scientific) and stored at -80°C. The quantitative RT-PCR reactions were done using Applied Biosystems™ TaqMan™ RNA-to-CT™ 1-Step Kit and Custom TaqMan Gene Expression Assays (Thermo Fisher Scientific). An mRNA standard curve was used to determine absolute mRNA concentrations for each of the samples. The mRNA standards were prepared as follows: separate CFPS reactions for 5 nM of plasmids (P70a-S28, P70a-deGFP, and P70a-deGFP-ssrA) were carried out for 2 hours. Total RNA was extracted using the full reaction volume. This was followed by the removal of 16S and 23S rRNA using the MICROBExpress™ Bacterial mRNA Enrichment Kit (Life Technologies Corporation). Lastly, the MEGAclear™ Kit (Life Technologies Corporation) was used to further purify the mRNA. The mRNA concentrations were determined using the Qubit™ RNA assay kit (ThermoFisher Scientific). At least three technical replicates were performed for each standard. The concentration of cI-ssrA mRNA was quantified using the deGFP-ssrA mRNA standard. Green fluorescent protein (deGFP) fluorescence was measured using the Varioskan Lux plate reader at 488 nm (excitation) and 535 nm (emission). At the end of the CFPS run, 3 μL of the reac-

tion mixture was diluted in 33 μL phosphate buffered saline (PBS) and stored at -80°C. The fluorescence was measured in triplicate with 10 μL each of this mixture. For all measurements, at least three biological replicates were performed.

3.3.3 Synthetic circuit architecture

The two genetic circuits ($C1$ and $C2$) used in this study were based upon the bacterial sigma factor regulatory system (Fig. 3.1). Sigma factor 70 (σ_{70}), endogenously present in the extract, was the primary driver of each circuit. In $C1$, σ_{70} induced green fluorescent protein (deGFP) expression was explored in the absence of additional regulators or protein degradation (Fig. 3.1A). In $C2$, σ_{70} induced the expression of sigma factor 28 (σ_{28}) and deGFP-ssrA (Fig. 3.1B). Sigma 28 induced the expression of the lambda phage repressor protein cI-ssrA, which was under the σ_{28} responsive P28 promoter. The cI-ssrA protein repressed the P70a promoter, thereby down-regulating σ_{28} and deGFP-ssrA transcription [127]. Simultaneously, the C-terminal ssrA degradation tags present on the deGFP and cI proteins were recognized by the endogenous ClpXP protease system in the cell free extract, thereby promoting the degradation of these proteins into peptide fragments [52, 59]. In addition, messenger RNAs (mRNAs) were always subject to degradation due to the presence of degradation enzymes in the extract [103, 59]. Taken together, the interactions of the components manifested in an accumulation of deGFP protein for $C1$, and a pulse signal of deGFP-ssrA in $C2$. Studying $C1$ allowed us to estimate parameters governing the interaction of σ_{70} with the P70a promoter. Whereas, the $C2$ allowed us to characterize the interaction of σ_{28} with the P28 promoter, the strength of the transcriptional repression by cI-ssrA, and the kinetics of protein degradation by

the endogenous ClpXP protease system. Finally, both circuits tested the effective model formulation for the transcription and translation rates.

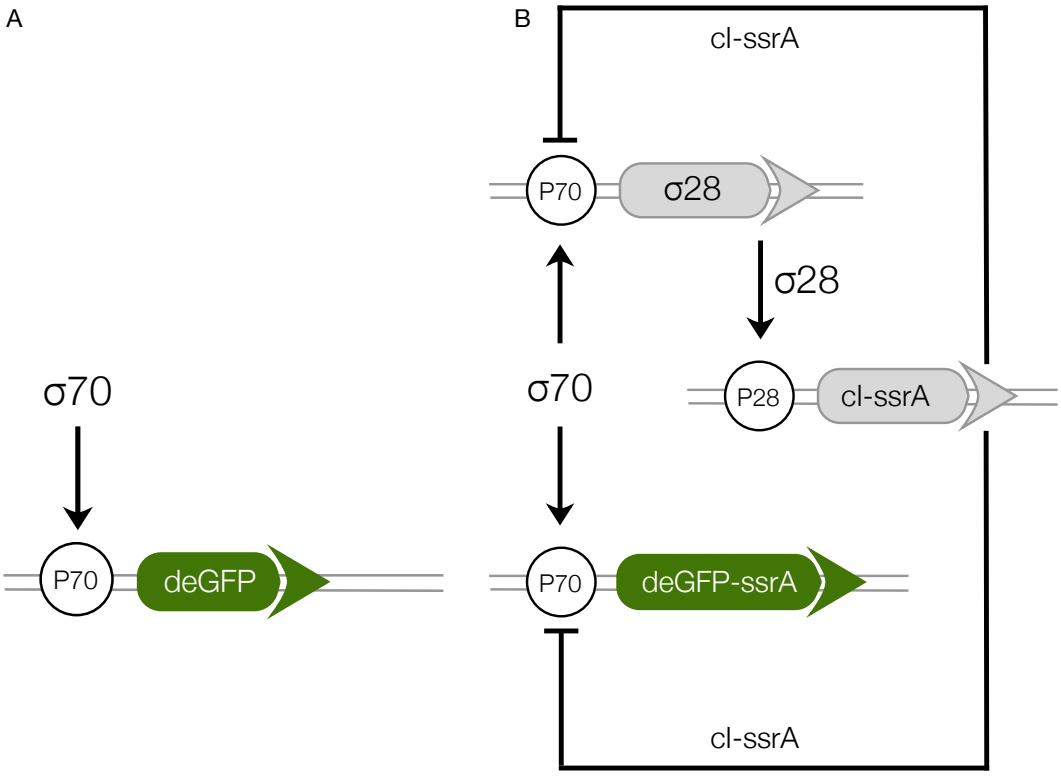


Figure 3.1: Schematic of the cell free gene expression circuits used in this study. **A**: Sigma factor 70 (σ_{70}) induced expression of deGFP. **B**: The circuit components encode for a negative feedback loop motif. Sigma factor 28 and deGFP-ssrA genes on the P70a promoters are expressed first because of the endogenous presence of sigma 70 factor in the extract. Sigma factor 28, once expressed, induces the P28a promoter, turning on the expression of the cI-ssrA gene which represses the P70a promoter. The circuit is modified from a previous study [59] by including an ssrA degradation tag on the cI gene.

3.3.4 Formulation and solution of model equations

Consider a cell free synthetic circuit composed of the genes $\mathcal{G} = 1, 2, \ldots, \mathcal{N}$. Each gene in the circuit is described by two differential equations, one for mRNA (m_j)

and a second for the corresponding protein (p_j):

$$\dot{m}_j = r_{X,j} u_j(\ldots) - \theta_{m,j} m_j \quad j = 1, 2, \ldots, \mathcal{N} \tag{3.1}$$

$$\dot{p}_j = r_{L,j} w_j(\ldots) - \theta_{p,j} p_j \tag{3.2}$$

The term $r_{X,j} u_j(\ldots)$ in the mRNA balance, which denotes the regulated rate of transcription for gene j, is the product of a kinetic limit $r_{X,j}$ (nM h^{-1}) and a transcription control function $0 \leq u_j(\ldots) \leq 1$ (dimensionless). Similarly, the rate of translation of mRNA j, denoted by $r_{L,j} w_j$, is also the product of the kinetic limit of translation (nM h^{-1}) and a translational control term $0 \leq w_j(\ldots) \leq 1$ (dimensionless). Lastly, $\theta_{\star,j}$ denotes the first-order rate constant (h^{-1}) governing degradation of protein and mRNA. The model equations, encoded in the Julia programming language [15], were automatically generated using the JuGRN tool [230]. The model equations were solved numerically using the `Rosenbrock23` routine of the `DifferentialEquations.jl` package [176].

Transcription and translation kinetic limits.

The key idea behind the transcription and translation kinetic limit expressions is that the polymerase (or ribosome) acts as a pseudo-enzyme; it binds a gene (or message), reads the gene (or message), and then dissociates. Thus, we used a strategy similar to classical enzyme kinetics to derive expressions for $r_{X,j}$ (or $r_{L,j}$); we proposed a set of elementary reactions for transcription and translation, one of which we assumed was rate limiting, and then invoked the pseudo state assumption for each intermediate complex to develop the overall rate expression. Following this approach, the details of the derivation of $r_{X,j}$ (or $r_{L,j}$) are given in the supplemental materials. The transcription kinetic limit $r_{X,j}$ is given

by:

$$r_{X,j} = V_{X,j}^{max}\left(\frac{\mathcal{G}_j}{\tau_{X,j}K_{X,j} + \left(1+\tau_{X,j}\right)\mathcal{G}_j + O_{X,j}}\right) \tag{3.3}$$

where $V_{X,j}^{max}$ denotes the maximum transcription rate (nM/h) of gene j, $\mathcal{G}_j$ denotes the concentration of gene j (nmol/L), $K_{X,j}$ denotes the saturation constant for transcription of gene j (nmol/L), $\tau_{X,j}$ denotes the time constant for transcription (dimensionless) and:

$$O_{X,j} = \sum_{i=1,j}^{N} \frac{K_{X,j}\tau_{X,j}}{K_{X,i}\tau_{X,i}}\left(1+\tau_{X,i}\right)\mathcal{G}_i \tag{3.4}$$

denotes the coupling of the transcription of gene j with the other genes in the system through competition for RNA polymerase.

In a similar way, we developed an expression for the translational kinetic limit:

$$r_{L,j} = V_{L,j}^{max}\left(\frac{m_j}{\tau_{L,j}K_{L,j} + \left(1+\tau_{L,j}\right)m_j + O_{L,j}}\right) \tag{3.5}$$

where $V_{L,j}^{max}$ denotes the maximum translation rate (nM/h), $K_{L,j}$ denotes the saturation constant for translation of mRNA message j (nmol/L), $\tau_{L,j}$ denotes the time constant for translation of message j (dimensionless) and:

$$O_{L,j} = \sum_{i=1,j}^{N} \frac{K_{L,j}\tau_{L,j}}{K_{L,i}\tau_{L,i}}\left(1+\tau_{L,i}\right)m_i \tag{3.6}$$

describes the coupling of the translation of mRNA j with other messages in the system because of kinetic competition for available ribosomes. The saturation and time constants for each case (which are unknown and must be estimated from experimental data) are defined in the supplemental materials. Lastly, in this study, the $O_{X,j}$ and $O_{L,j}$ terms were neglected as both circuits had either only one, or a small number of genes.

The maximum transcription rate $V_{X,j}^{max}$ was formulated as:

$$V_{X,j}^{max} \equiv R_{X,T}\left(\frac{\dot{v}_X}{l_{G,j}}\right) \tag{3.7}$$

where $R_{X,T}$ denotes the total RNA polymerase concentration (nM), $\dot{v}_X$ denotes the RNA polymerase elongation rate (nt/h) and $l_{G,j}$ denotes the length of gene j in nucleotides (nt). Similarly, the maximum translation rate $V_{L,j}^{max}$ was formulated as:

$$V_{L,j}^{max} \equiv K_P R_{L,T}\left(\frac{\dot{v}_L}{l_{P,j}}\right) \tag{3.8}$$

where $R_{L,T}$ denotes the total ribosome pool, K_P denotes the polysome amplification constant, $\dot{v}_L$ denotes the ribosome elongation rate (amino acids per hour), and $l_{P,j}$ denotes the length of protein j (aa).

Control functions u and w.

Values of the control functions $u(\ldots)$ and $w(\ldots)$ describe the regulation of transcription and translation. Ackers et al., borrowed from statistical mechanics to recast the $u(\ldots)$ function as the probability that a system exists in a configuration which leads to expression [2]. The idea of recasting $u(\ldots)$ as the probability of expression was also developed (apparently independently) by Bailey and coworkers in a series of papers modeling the *lac* operon, see [115]. More recently, Moon and Voigt adapted a similar approach when modeling the expression of synthetic circuits in *E. coli* [143]. The $u(\ldots)$ function is formulated as:

$$u(\ldots)_j = \frac{\sum_{i\in\{\chi\}} W_i f_i(\ldots)}{\sum_{j\in C_j} W_j f_j(\ldots)} \tag{3.9}$$

where W_i (dimensionless) denotes the weight of configuration i, while $f_i(\cdots)$ (dimensionless) is a binding function (taken to be a hill-type function) describ-

ing the fraction of bound activator/inhibitor for configuration i. The summation in the numerator of Eqn (3.9) is over the set of promoter configurations leading to expression (denoted as χ), while the summation in the numerator is over the set of all possible configurations for gene j (denoted as C_j). Thus, $u(\ldots)_j$ can be thought of as the fraction of all possible configurations that lead to expression. The weights W_i are related to the Gibbs energy of configuration i: $W_i = \exp(-\Delta G_i/RT)$ where ΔG_i denotes the molar Gibbs energy for configuration i (kJ/mol), R denotes the ideal gas constant (kJ mol^{-1} K^{-1}), and T denotes the system temperature (Kelvin) [2]. The value of the binding function depends on the concentrations of the different transcriptional elements and their dissociation constants. The temporal evolution of u, therefore, is tied to the dynamics of its transcriptional elements, and its value lies between 0 and 1. In the case of circuit $C1$, u did not vary during the course of the reaction because the concentration of its activator, σ_{70}, was fixed. For this case, u approximately equalled 0.95. However, in the second circuit, $C2$, u varied with time because of the change in levels of σ_{28} and cI-ssrA proteins.

We accounted for the experimentally observed loss of translational activity through the translational control function $w(\ldots)$. Loss of translational activity could be a function of many factors, including depletion of metabolic resources. However, in this study, we modeled the loss of translational activity as an exponential decay process with half-life $\tau_{L,1/2}$:

$$\dot{\epsilon} = -\left(\frac{0.693}{\tau_{L,1/2}}\right)\epsilon \tag{3.10}$$

where ϵ denotes the fraction of remaining translational activity. Initially we assumed translation to be fully active, $\epsilon(0) = 1$. Solving equation (10) yields $\epsilon(t) = \exp(-0.693 \cdot t/\tau_{L,1/2})$. Over time, as the cell free reaction progressed, the translational activity decreased with a half-life $\tau_{L,1/2}$ which was estimated from

experimental data. The translational control variable was then given by $w_i = \epsilon$ for all translation processes.

3.3.5 Estimation of model parameters

Model parameters were estimated from published studies, were specified by experimental conditions (Table 3.1) or were estimated by minimizing the squared difference between model simulations and messenger RNA (mRNA), or protein measurements generated in this study. For the P70-deGFP model ($C1$), 11 parameters were estimated, while 33 parameters were estimated for the negative feedback model ($C2$).

The minimization problem to estimate the unknown model parameters was structured as a multiobjective optimization problem in which each measured mRNA or protein trajectory was treated as a separate objective. The minimization problem was solved using the Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) [13]. JuPOETs is a multiobjective optimization approach which integrates simulated annealing with Pareto optimality to estimate parameter values on or near the optimal tradeoff surface between N potentially competing objectives (squared difference between model simulations and experimental measurements). JuPOETs minimized a problem of the form:

$$\min_k \mathcal{E}_j = \sum_{i=1}^{\mathcal{T}_j} \left(\hat{\mathcal{M}}_{ij} - x_{ij}(k) \right)^2 \qquad j = 1, 2, \ldots, N \tag{3.11}$$

subject to

$$\dot{x} = f(x,k) \tag{3.12}$$

$$\mathcal{L} \leq k \leq \mathcal{U} \tag{3.13}$$

$$x(t_o) = x_o \tag{3.14}$$

where Eqn (3.12) denotes the model equations, Eqn (3.13) denotes the parameter bounds, and Eqn (3.14) denotes the initial conditions. The objective function(s) $\mathcal{E}_j$ measured the squared difference between model simulations and experiment j (either a protein or mRNA trajectory). The symbol $\hat{\mathcal{M}}_{ij}$ denotes an experimental observation at time index i from experiment j, while the symbol x_{ij} denotes the model simulation output at time index i from experiment j. The quantity i denotes the sampled time-index and $\mathcal{T}_j$ denotes the number of time points for experiment j. For the P70-deGFP model ($C1$), $\mathcal{E}_1$ corresponded to mRNA deGFP, while $\mathcal{E}_2$ corresponded to the deGFP protein concentration. On the other hand, for the negative feedback model ($C2$), $\mathcal{E}_1$ corresponded to mRNA deGFP-ssrA, $\mathcal{E}_2$ to mRNA σ_{28}, $\mathcal{E}_3$ to mRNA cI-ssrA and $\mathcal{E}_4$ to the deGFP-ssrA protein concentration. Lastly, we penalized accumulation of the cI-ssrA protein (unmeasured) reaching unrealistically high levels with a term of the form: $\mathcal{E}_5 = C \times \max(0, x_{cI} - U_{cI})$ where C denotes a penalty parameter ($C = 1{\times}10^5$), x_{cI} denotes the maximum simulated cI-ssrA protein concentration, and U_{cI} denotes a concentration upper bound ($U_{cI} = 100\mu M$). This bound was chosen to be approximately five-fold higher than the protein levels observed in an uninhibited circuit (e.g., $C1$).

The lower and upper bounds for unknown model parameters were established from previously published studies, or from previous model analysis; parameter values estimated for the P70-deGFP model were also used to establish

ranges for the negative feedback model. JuPOETs searched over ΔG_i, K_L and $\tau_{L,1/2}$ values directly, while other unknown parameter values took the form of corrections to order of magnitude characteristic literature estimates. For example, we set the mRNA degradation rate constant (θ_m) to a characteristic value taken from literature. Then, the degradation constant for any particular mRNA was represented as: $\theta_{m,i} = \alpha_i \theta_m$, where α_i was an unknown (but bounded) modifier. In this way, we guaranteed the parameter search (and the resulting estimated parameters) were within a specified range of literature values. We used this procedure for all degradation constants (both mRNA and protein) and all time constants (for both transcription and translation). The baseline parameter values are given in Table 3.1. JuPOETs was run for 20 generations for both models, and all parameters sets with Pareto rank less than or equal to two were collected for each generation. The JuPOETs parameter estimation routine is encoded in the `sa_poets_estimate.jl` script in the model repositories.

JuPOETs uses a simulated annealing approach to generate candidate parameter solutions whose Pareto rank is then evaluated; ranks below a threshold are kept while higher rank solutions are discarded. Thus, all the advantages (and disadvantages) associated with simulated annealing have been inherited by JuPOETs; for example, the time required to generate a family of low rank solutions will be significantly longer than a derivative based approach. Beyond these specific performance issues, a unique pathology of JuPOETs is the use of Pareto rank as a surrogate for training error. JuPOETs attempts to find low rank solutions, but rank is a relative measure of the quality of a solution. Thus, during the early iterations, low rank solutions often have large errors. As the iteration count increases the approach tends to find low error solutions with low rank, however, for complex models the rate of convergence to these low rank

low error solutions is slow. To address this concern, we periodically switch to single objective mode where we minimize the total training error (summation of all objective functions) instead of finding low rank solutions. The best solutions from single objective mode can then be used to restart the multi-objective calculation. This hybrid approach, which was used in this study, has previously been shown to increase the rate of finding low rank and low error solutions, see [13].

3.3.6 Morris sensitivity analysis

Morris sensitivity analysis was used to understand which model parameters were sensitive [147]. The Morris method is a global method that computes an elementary effect value for each parameter by sampling a model performance function, in this case the area under the curve for each model species in their respective timeplots, over a range of values for each parameter; the mean of elementary effects measures the direct effect of a particular parameter on the performance function, while the variance of each elementary effect indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests connectedness or nonlinearity). The Morris sensitivity coefficients were computed using the `DiffEqSensitivity.jl` package [176]. The parameter ranges were established by calculating the minimum and the maximum value for each parameter in the parameter ensemble generated by JuPOETs. Each range was then subdivided into 10,000 samples for the sensitivity calculation. Elementary effect values were then calculated one at a time by measuring the perturbation in the AUC on changing one parameter, where the AUC was calculated by solving the set of ODEs for each change. In order

to calculate the mean and variance, the top 1000 perturbations with the highest spread in parameter values were used. In total, the model was evaluated $10000n$ times, where n is the number of parameters varied. The Morris sensitivity coefficients are calculated using the `compute_sensitivity_coefficients.jl` script in the model repositories.

3.4 Results

3.4.1 Modeling and analysis of the $C1$ circuit

The effective biophysical transcription and translation model captured σ_{70} induced deGFP expression at the mRNA and protein level within the experimental error for $C1$ (Fig. 3.2). JuPOETs produced an ensemble (N = 140) of the 11 unknown model parameters which captured the transcription of mRNA (Fig. 3.2A) and the translation of deGFP protein (Fig. 3.2B). The mean and standard deviation of key parameters is given in Table 3.2. The deGFP mRNA reached its steady state concentration of approximately 580 nM within two hours, and stayed at this level for the remainder of the reaction. Thus, the cell free reaction maintained continuous transcriptional activity with an average mRNA lifetime of 27 minutes; Garamella et al [59] reported a similar lifetime of 17 - 18 min. On the other hand, deGFP protein concentration increased more slowly, and began to saturate between 8 to 10 hr at approximately 15 μM. Given there was negligible protein degradation (the mean deGFP half-life was estimated to be ~11 days, which was similar to the value of 6 days estimated by Horvath et al, albeit in a different cell free system [84]). The saturating protein concentration

suggested that the translational capacity of the cell free system decreased over the course of the reaction. The decrease in translational capacity, which could stem from several sources, was captured in the simulations using a monotonically decreasing translation capacity state variable ϵ, and the translational control variable $w(\ldots)$. In particular, the mean half-life of translational capacity was estimated to be $\tau_{L,1/2} \sim 4$ h in the $C1$ experiments. Taken together, JuPOETs produced an ensemble of model parameters that captured the experimental training data. Next, we considered which $C1$ model parameters were important to the model performance using Morris sensitivity analysis, a global sensitivity analysis method.

The importance of $C1$ model parameters was quantified using Morris sensitivity analysis (Fig. 3.2B). The Morris method computes the influence of each parameter, known as the elementary effect, on a model performance function. The mean of elementary effects measures the direct effect of a particular parameter, while the variance indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests nonlinearity). The performance function was defined as the integrated area under the curve (AUC) for each mRNA and protein species in their respective timeplots, calculated for each parameter value range. The Morris sensitivity measures (mean and variance) were binned into categories based upon their relative magnitudes, from no influence (white) to high influence (black). Only four parameters (translation saturation coefficient K_L, translational capacity half-life $\tau_{L,1/2}$, translation time constant, and protein degradation constant) influenced the protein level. On the other hand, six parameters influenced both mRNA and protein abundance; all six of these parameters were either directly or indirectly associated with transcription. Thus, these parameters influenced the

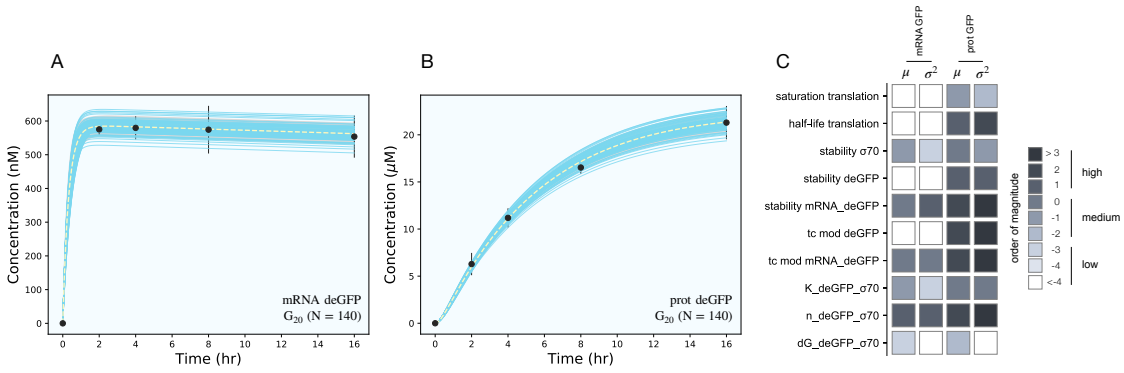


Figure 3.2: Model simulations versus experimental measurements for σ_{70} induced deGFP expression. **A**: Simulated and measured deGFP mRNA concentration versus time using the small circuit G_{20} ensemble (N = 140). **B**: Simulated and measured deGFP protein concentration versus time the small circuit G_{20} ensemble (N = 140). **C**: Global sensitivity analysis of the P70-deGFP circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble. Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 140 parameter sets that were rank two or below. The parameter ensemble was used to calculate the mean (dashed line) and the 95% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates.

production or stability of mRNA which in turn influenced the protein level. In particular, the mRNA degradation constant and the cooperativity of σ_{70} in the P70a promoter function had the largest direct effect and variance. Surprisingly, the ΔG of σ_{70}/RNAP/promoter configuration was the least influential of the six parameters and had a small elementary effect variance. Taken together, Morris sensitivity analysis of the $C1$ model parameters highlighted the hierarchical structure of the transcriptional and translational model, suggesting experimentally tunable parameters such as mRNA stability were globally important. Next, we used the ensemble of P70a, time constant and degradation parameters esti-

mated for $C1$ to constrain the analysis of $C2$.

3.4.2 Modeling and analysis of the $C2$ circuit

The effective biophysical transcription and translation model captured the deGFP-ssrA expression dynamics in the negative feedback circuit $C2$ (Fig. 3.3A). JuPOETs produced an ensemble (N = 498) of the 33 unknown model parameters which captured transcription and translation dynamics for σ_{28}, cI-ssrA and deGFP-ssrA. The mean and standard deviation of key parameters is given in Table 3.3. Compared with the estimated parameters for $C1$, the $C2$ model had almost a two fold change in the half life of translation and the translation saturation coefficient. Similarly, there were variations in the values of the transcription and translation time constants for the two systems. However, for both circuits, the small values of the transcription and translation time constants qualitatively suggested elongation limited reactions; the exception was σ_{28} translation which was closer to initiation limited. Unlike $C1$, the mRNA expression pattern for σ_{28} and deGFP-ssrA both showed an initial spike, to a concentration similar with the previous pseudo steady state, before the cI-ssrA regulator protein could be expressed. However, once cI-ssrA began to accumulate, the concentrations of the regulated mRNAs dropped by approximately an order of magnitude compared with the unregulated case. Again, as shown with $C1$, the regulated mRNA concentrations reached an approximate steady-state. This further confirmed continuous transcription and mRNA degradation throughout the cell free reaction. The mean estimated mRNA lifetime for cI-ssrA and deGFP were similar (approximately 16 min), while the degradation of σ_{28} mRNA was predicted to be slower (mean mRNA lifetime was estimated

to be approximately 30 min). Lastly, the mean peak degradation rate for GFP was approximately 47 nM/min, while the mean peak cI-ssrA degradation rate was predicted to be approximately 63 nM/min; both of these degradation rate estimates were consistent with the previously reported range of 15 nM/min - 150 nM/min measured by Garamella et al [59].

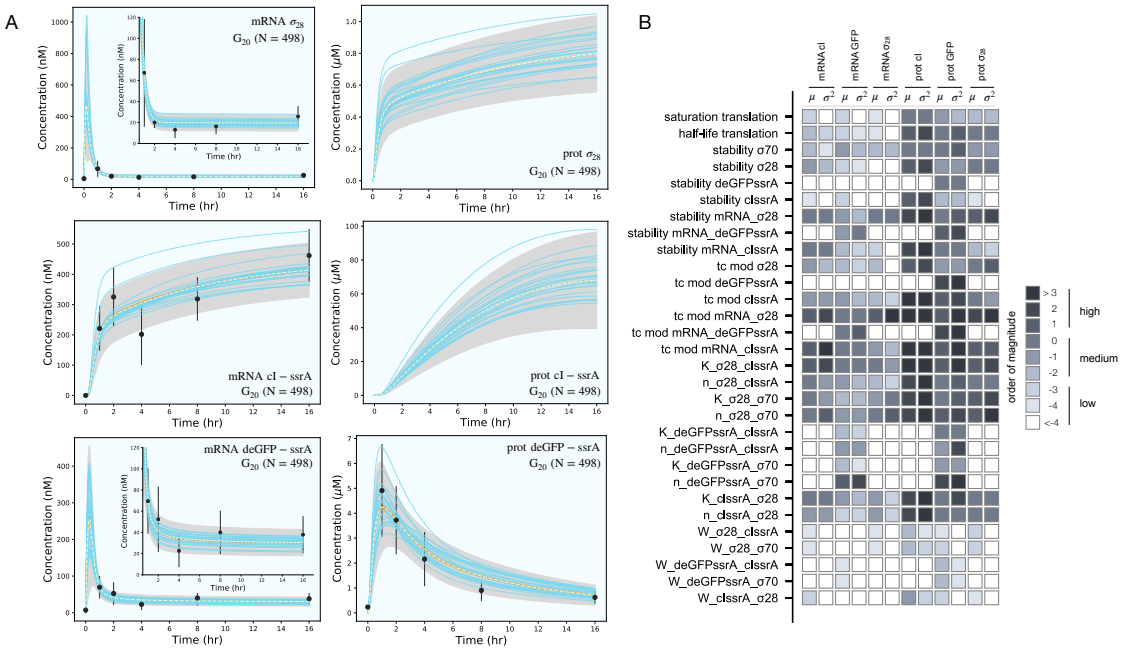


Figure 3.3: Model simulations versus experimental measurements for the negative feedback deGFP-ssrA circuit. **A**: Model simulations of the negative feedback deGFP-ssrA circuit using the G_{20} ensemble (N = 498). Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 489 parameter sets (rank two or below). The parameter ensemble was used to calculate the mean (dashed line) and the 99% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are also shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates. **B**: Global sensitivity analysis of the negative feedback deGFP-ssrA circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble.

The secondary effect of cI-ssrA repression was visible in the cI-ssrA mRNA

expression pattern. The expression of cI-ssrA was induced by σ_{28}, however, σ_{28} expression was repressed by cI-ssrA, thereby completing a negative feedback loop. Initially, before appreciable levels of cI-ssrA had been translated, the cI-ssrA transcription rate was maximum (approximately 200 nM/h). However, the transcription rate decreased to approximately 12 nM/h after two hours and remained constant for the remainder of the cell free reaction. Similarly, transcription rates for σ_{28} (approximately 1200 nM/h) and deGFP-ssrA (approximately 750 nM/h) were initially at a maximum due to the presence of endogenous σ_{70}, but then quickly dropped as cI protein accumulated. Protein synthesis followed a similar trend, with the translation rates for σ_{28} and deGFP-ssrA initially present at their maximum values before quickly dropping. After one hour, deGFP levels reached a peak and decayed due to the ClpXP-mediated degradation, whereas σ_{28} protein levels continued to slowly rise at a steady rate (approximately 15 nM/h). The $C2$ model also predicted the expected lag present during the initial phase of cI-ssrA protein synthesis due to the need for σ_{28} protein to reach appreciable levels. Moreover, the combination of high cI-ssrA mRNA abundance (expressed because σ_{28} does not have a degradation tag) and ClpXP-mediated degradation led to the saturation of the cI-ssrA protein concentration. However, the cI-ssrA protein concentration could not be verified because we did not have an experimental measurement for this species. Taken together, the effective model simulated cell free expression dynamics for $C2$. Next, we considered which $C2$ model parameters were important using Morris sensitivity analysis.

Morris sensitivity analysis of the negative feedback circuit $C2$ stratified the parameters into locally and globally important groups (Fig. 3.3B). The influence of 33 parameters was computed using the AUC of each mRNA and pro-

tein species as the performance function. The Morris sensitivity metrics (mean and variance) were binned into categories based upon their relative magnitudes, from no influence (white) to high influence (black). Some parameters affected only their respective mRNA or protein target, whereas others had widespread effects. For example, the time constant (tc) modifiers, stability of deGFP-ssrA protein and mRNA, and the binding dissociation constant (K) and cooperativity parameter (n) of cI-ssrA and σ_{70} for the deGFP-ssrA promoter affected only the values of deGFP-ssrA protein and mRNA. On the other hand, the tc, stability, K and n parameters for σ_{70}, σ_{28}, or cI-ssrA influenced mRNA and protein expression globally. The σ_{70} and σ_{28} proteins acted as inducers or repressors for more than one gene product: σ_{70} induced both deGFP-ssrA and σ_{28}, and cI-ssrA protein repressed both of these genes. Degradation constants (denoted as stability) affected the half-lives of the transcribed messages or the translated proteins in the mixture, while the time constant modifiers changed the time required to form the open gene complex (or translationally active complex). Dissociation and cooperativity constants affected the binding interactions of the inducer (or repressor in the case of cI-ssrA) in the promoter control function. Varying these parameters, therefore, had a strong effect on their respective targets. Similarly, the translation saturation and its half-life, which captured the depletion in the translation activity over the course of the reaction, not only affected protein levels but also mRNA levels. This is because these parameters tuned the rate of formation of cI-ssrA, which in turn affected the mRNA levels of its gene targets. Given that cI-ssrA was the main regulator (repressor) of the circuit, the parameters that dictated the levels of cI-ssrA mRNA and protein had a global effect. We also observed high sensitivity variance for several parameters, in particular involving cI-ssrA. For example, the time constant modifiers for cI-ssrA

mRNA and protein had a two-pronged effect. On the one hand, they positively influenced the transcription/translation rates of the gene and mRNA products, directly increasing the cI-ssrA protein. On the other hand, increased cI-ssrA expression reduced the level of σ_{28}, in turn reducing the cI-ssrA levels over time. Taken together, Morris sensitivity analysis of the $C2$ model stratified that parameters into local and globally important groups, with the parameters governing the synthesis rates of the cI-ssrA mRNA and protein being globally important. The sensitivity analysis also gave insight into the organization of the circuit, suggesting cI to be highly connected within the circuit.

3.5 Discussion

In this study, we developed an effective biophysical modeling approach to simulate transcription (TX) and translation (TL) processes occurring in a cell free system. We tested this approach by simulating the dynamics of two cell free synthetic circuits ($C1$ and $C2$). The model formulation, and parameter values were mechanistic and largely derived from literature. For example, characteristic values for τ_X and K_X, the time and saturation constants for transcription, were approximated from *in vitro* experiments using an abortive initiation assay [136]. The RNAP and ribosome elongation rates were taken from Garamella et al. [59], while other parameters were estimated from BioNumbers [140]. Similarly, the weights appearing in the transcription control function $u(...)$ were based upon the Gibbs energies of the respective promoter configurations, while the form of the transcriptional control functions was derived from a statistical mechanical treatment of promoter activity [2, 115, 143]. However, there were parameters that were not available from literature; in these cases multiobjective

optimization was used to estimate these parameters from mRNA and protein measurements. For $C1$, sigma factor 70 (σ_{70}) induced expression of green fluorescent protein (deGFP), the time constants, degradation rates, and other parameters governing deGFP expression were estimated from measurements of deGFP mRNA and protein. These estimates were then used to constrain the parameter search for $C2$, which involved deGFP expression subject to negative feedback and programmed protein degradation. We estimated which model parameters were important to the performance of $C1$ and $C2$ using Morris sensitivity analysis. Sensitivity analysis results for $C1$ were expected; the time constant for transcription, the stability of the deGFP message and the cooperativity of σ_{70} were all important parameters. On the other hand, the sensitivity analysis results for $C2$ were more nuanced, with parameters (and associated species) being stratified into locally and globally important groups; the performance of $C2$ was most sensitive to the parameters controlling the cI-ssrA mRNA and protein abundance.

The effective TX/TL modeling approach described here has several potential applications. For example, a challenge of *in vivo* constraint based modeling is the description of gene expression [36]. Boolean and probabilistic approaches [34, 35, 27] have been developed to address this challenge. However, the transcriptional state of a boolean gene is either on or off based on the state of its regulators, thus, gene expression is coarse-grained. The current modeling approach could be an interesting mechanistic alternative to the boolean framework that utilizes a continuous description of gene expression dynamics and transcriptional regulation. In particular, the rules encoding typical boolean gene expression networks are easily translatable into the rational promoter functions described here, however, the estimation of the parameters appearing in these

promoter functions, especially in an *in vivo* context, remains an open question. Another application could be the extension of the current model to other prokaryotic or eukaryotic systems with a few changes. For example, in order to adopt it for an *in vivo* system, the dilution of resources due to growth (proportional to the cellular doubling time) would be added as a first order term to the mRNA and protein balance equations. Additionally, the competition for RNAP and ribosomes, denoted respectively as $O_{X,j}$ and $O_{L,j}$ in the study, and assumed to be negligible due to the presence of only three genes in the system, would need to be taken into account; this term would serve to change the rates of transcription and translation of the added genes because of the presence of a large amount of endogenous genes in the *in vivo* system. Moreover, characteristic literature-based parameter values would be different for cellular processes compared to the *in vitro* ones used in this study, and they would thus need to be adjusted accordingly. For the case of a mammalian or a yeast *in vivo* system, a few more changes to the current model are necessary because the mechanistic processes of gene expression and regulation are different in these two types of systems. For example, a key difference present in eukaryotes is the addition of an intron splicing step during the synthesis of a mature mRNA from a pre-mRNA. In addition, the gene regulation mechanisms are vast and composed of numerous elements in eukaryotes. Finally, especially in *in vivo* systems, addition of exogenous genes often leads to a tug-of-war of carbon and energy resources between cellular growth processes and the expression of these genes, driving cellular resources away from the latter. Synthetic biology studies often neglect the role that metabolism plays in the expression of synthetic circuits. Ultimately, metabolism is centrally important to the operation of any synthetic circuit; gene expression is strongly dependent upon the metabolic resources pro-

vided by catabolic processes. It is imperative that this metabolic burden by the addition of exogenous genes be incorporated in the *in vivo* model description to accurately capture the expression behavior. We have recently started to explore this question by integrating effective transcription and translation models with metabolic networks in cell free reactions e.g., [84, 237], and also developing experimental tools to measure metabolite concentrations in cell free systems [235]. However, these previous transcriptional and translations models (and similar precursor models simulating eukaryotic processes [67, 216]) were less developed than those presented here. Taken together, the effective modeling approach described here can potentially be used to simulate transcription and translation processes in a variety of applications.

There have been many studies looking into oscillatory and other dynamic behavior of synthetic circuits, see Prangemeier et al [172]. A negative feedback loop, such as the one explored here, has the potential to give rise to oscillations. Yelleswarapu et al carried out TX/TL reactions, with a circuit similar to $C2$, in both batch and continuous conditions [249]. Similar to our study, no oscillations were observed in the batch reaction. However, oscillations were observed in the continuous reaction. There are several possible reasons why no oscillations were seen in our (or the Yelleswarapu et al) batch study; as it was carried out in batch, dilution of the expressed protein or mRNA species due to an inlet feed was not possible. Thus, mRNA species reached a pseudo steady state (after approximately 2 hours) because of ribonuclease degradation [61]. On the other hand, in general protein species were not at steady-state; only proteins tagged with a ssrA tag were able to be degraded by the ClpXP system, thereby allowing a steady-state. Thus, the batch system likely evolved dynamically through a set of concentration profiles that were not consistent with oscillations.

The effective TX/TL model described the experimental mRNA and protein training data. However, there were several important questions to be addressed by future studies. First, the model formulation described the data, but did not predict dynamics outside of the training set. If this approach is to be useful to the synthetic biology community, or more broadly as an effective biophysical technique to model *in vivo* gene expression dynamics for applications such as regulatory flux balance analysis, we need to have confidence that the modeling approach is predictive. Thus, while we have established a descriptive model, we have to yet to establish a predictive model. Next, there were several technical or mechanistic questions that should be explored further. For example, cI-ssrA represses the activity of the P70a promoter via interaction with its OR2 and OR1 operator sites; in this study we considered only a single operator site suggesting that we potentially underestimated the potency of cI repression in the deGFP and σ_{28} promoter functions, see the multiplication rule [124]. Further, we used a first order approximation of ClpXP mediated protein degradation, while Garamella et al. [59] described this degradation as zero order. Similarly, we did not establish the concentration of ClpXP in the commercially available cell free reaction mixture. The levels of this protein complex could be an important factor controlling protein degradation. Next, we should compare the current modeling approach, and the values estimated for the model parameters, with the study of Marshall and Noireaux [128]. For example, one of the potential limitations of the current study (that was addressed by Marshall and Noireaux [128]) is that we did not consider a separate species for dark GFP. In our previous RNA circuit modeling [86], we did include this term, but failed to do so here. We expect inclusion of a dark versus light GFP species could influence the values for the estimated parameters, particularly the translation time constants. However,

previous reports suggested the *in vitro* maturation time of deGFP was approximately 8 min [190], much faster than the typical maturation times for GFP of one hour *in vivo* [199, 91]. Thus, the impact of including a dark versus light GFP species may not be worth the increased model complexity. Lastly, we should validate the values estimated for the binding function parameters and the promoter configuration free energies. Maeda et al measured the binding affinities of the seven *E. coli* σ factors with RNAP [126]; while not directly comparable, these measurements give an order of magnitude characteristic value for the dissociation constants appearing in the promoter binding functions. Further, there have been several studies that have quantified the binding energies of promoter configurations e.g., [2, 21, 214, 215]. A perfunctory inspection of the values estimated in this study suggested our estimated free energy values, while the same order of magnitude as previous studies in many cases, did have values that were off by a factor of up to an order of magnitude compared with literature (albeit for different promoters). In particular, the positive Gibbs energy estimated for free RNAP binding leading to transcription was likely too large, while the magnitude of other values such as the energy of cI repression of σ_{28} expression was likely too small. Thus, these other studies could serve as a basis to validate our estimates, and perhaps more importantly constrain the parameter search space for future studies.

Table 3.1: Characteristic parameters for TX/TL model equations. Key to references used in the table: (a) [59], (b) [227], (c) set by experiment, (d) [106], (e) estimated in this study, (f) [153], (g) [70], (h) calculated from plasmid sequence, (i) [136], (f) [237]

Description	**Parameter**	**Value**	**Units**	**Reference**
RNA polymerase concentration	$R_{X,T}$	0.06-0.075	μM	*a*
Ribosome concentration	$R_{L,T}$	< 2.3	μM	*a,b*
σ_{70} concentration	σ_{70}	<35	nM	*a*
initial σ_{28} concentration	σ_{28}	<20	nM	*a*
Transcription elongation rate	$\dot{v}_X$	12-30	nt/s	*a,d*
Translation elongation rate	$\dot{v}_L$	1-2	aa/s	*a,b*
Transcription saturation coefficient	K_X	0.036	μM	*i*
Polysome amplification constant	K_P	10.0	constant	*e*
Transcription initiation time	k_I^X	22	s	*i*
Translation initiation time	k_I^L	1.5	s	*e*
Default mRNA degradation coefficient	θ_m	3.75	h^{-1}	*a*
Default protein degradation coefficient	θ_p	0.462-1.89	h^{-1}	*f,g*
Gene concentration σ_{28}		1.5	nM	*c*
Gene concentration cI-ssrA		1.0	nM	*c*
Gene concentration deGFP-ssrA		8.0	nM	*c*
Gene length σ_{28}		811	nt	*h*
Gene length cI-ssrA		850	nt	*h*
Gene length deGFP-ssrA		782	nt	*h*
Protein length σ_{28}		240	aa	*h*
Protein length cI-ssrA		248	aa	*h*
Protein length deGFP-ssrA		237	aa	*h*

Table 3.2: Estimated parameter values for the P70-deGFP model ($C1$). The mean and standard deviation of each parameter value was calculated over the ensemble of parameter sets meeting the rank selection criteria (N = 139).

Description	**Parameter**	**Value** ($\mu \pm \sigma$)	**Units**
Translation saturation coefficient	K_L	483.13 ± 10.10	μM
Half-life translation	$\tau_{L,1/2}$	4.03 ± 0.031	h^{-1}
Time constants			
deGFP transcription	$\tau_{X,GFP}$	0.61 ± 0.04	dimensionless
deGFP translation	$\tau_{L,GFP}$	0.16 ± 0.003	dimensionless
mRNA and protein half-life			
mRNA deGFP	$\ln(2)/\theta_{m,GFP}$	13.5 ± 2.47	min
Protein deGFP	$\ln(2)/\theta_{p,GFP}$	10.86 ± 0.78	days
Protein σ_{70}	$\ln(2)/\theta_{p,\sigma_{70}}$	3.65 ± 0.17	days
Free energies			
RNAP + deGFP gene	$\Delta G_{GFP,RX}$	28.82 ± 1.75	kJ mol^{-1}
RNAP + σ_{70} + deGFP gene	$\Delta G_{GFP,\sigma_{70}}$	-20.38 ± 1.91	kJ mol^{-1}
Binding parameters			
Hill coefficient deGFP gene + σ_{70}	$n_{GFP,\sigma_{70}}$	1.12 ± 0.06	dimensionless
Dissociation constant deGFP gene + σ_{70}	$K_{GFP,\sigma_{70}}$	24.19 ± 2.18	μM

Table 3.3: Estimated parameter values for the negative feedback circuit ($C2$). The mean and standard deviation for each parameter was calculated over the ensemble of parameter sets (N = 498).

Description	**Parameter**	**Value** ($\mu \pm \sigma$)	**Units**
Translation saturation coefficient	K_L	253.75 ± 14.12	μM
Half-life translation	$\tau_{L,1/2}$	8.86 ± 0.85	h^{-1}
Time constants			
cI-ssrA transcription	$\tau_{X,cI}$	< 0.001	dimensionless
deGFP transcription	$\tau_{X,GFP}$	0.045 ± 0.003	dimensionless
σ_{28} transcription	$\tau_{X,\sigma_{28}}$	0.0018 ± 0.0003	dimensionless
cI-ssrA translation	$\tau_{L,cI}$	0.054 ± 0.004	dimensionless
deGFP translation	$\tau_{L,GFP}$	0.058 ± 0.007	dimensionless
σ_{28} translation	$\tau_{L,\sigma_{28}}$	1.1 ± 0.13	dimensionless
mRNA and protein half-life			
mRNA cI-ssrA	$\ln(2)/\theta_{m,cI}$	8.1 ± 0.60	min
mRNA deGFP	$\ln(2)/\theta_{m,GFP}$	7.74 ± 1.13	min
mRNA σ_{28}	$\ln(2)/\theta_{m,\sigma_{28}}$	14.96 ± 1.60	min
Protein cI-ssrA	$\ln(2)/\theta_{p,cI}$	0.46 ± 0.043	days
Protein deGFP-ssrA	$\ln(2)/\theta_{p,GFP}$	0.051 ± 0.002	days
Protein σ_{28}	$\ln(2)/\theta_{p,\sigma_{28}}$	7.65 ± 0.91	days
Protein σ_{70}	$\ln(2)/\theta_{p,\sigma_{70}}$	14.86 ± 2.30	days
Free energies			
RNAP + cI gene	$\Delta G_{cI,RX}$	46.57 ± 4.28	kJ mol^{-1}
RNAP + σ_{28} + cI gene	$\Delta G_{cI,\sigma_{28}}$	-1.10 ± 0.04	J mol^{-1}
RNAP + deGFP gene	$\Delta G_{GFP,RX}$	41.94 ± 1.80	kJ mol^{-1}
RNAP + σ_{70} + deGFP gene	$\Delta G_{GFP,\sigma_{70}}$	-27.67 ± 1.79	kJ mol^{-1}
RNAP + cI + deGFP gene	$\Delta G_{GFP,cI}$	-7.21 ± 1.14	kJ mol^{-1}
RNAP + σ_{28} gene	$\Delta G_{\sigma_{28},RX}$	46.67 ± 3.18	kJ mol^{-1}
RNAP + σ_{70} + σ_{28} gene	$\Delta G_{\sigma_{28},\sigma_{70}}$	-10.46 ± 1.15	kJ mol^{-1}
RNAP + cI + σ_{28} gene	$\Delta G_{\sigma_{28},cI}$	-12.89 ± 1.44	kJ mol^{-1}

Table 3 continued

Description	Parameter	Value ($\mu \pm \sigma$)	Units
Hill coefficients			
cI gene + σ_{28}	$n_{cI,\sigma_{28}}$	1.88 ± 0.28	dimensionless
deGFP gene + σ_{70}	$n_{GFP,\sigma_{70}}$	1.53 ± 0.14	dimensionless
deGFP gene + cI	$n_{GFP,cI}$	0.698 ± 0.133	dimensionless
σ_{28} gene + σ_{70}	$n_{\sigma_{28},\sigma_{70}}$	1.10 ± 0.10	dimensionless
σ_{28} gene + cI	$n_{\sigma_{28},cI}$	1.51 ± 0.25	dimensionless
Dissociation constants			
cI gene + σ_{28}	$K_{cI,\sigma_{28}}$	1.09 ± 0.088	μM
deGFP gene + σ_{70}	$K_{GFP,\sigma_{70}}$	86.87 ± 7.13	μM
deGFP gene + cI	$K_{GFP,cI}$	3.83 ± 0.41	μM
σ_{28} gene + σ_{70}	$K_{\sigma_{28},\sigma_{70}}$	1.35 ± 0.26	μM
σ_{28} gene + cI	$K_{\sigma_{28},cI}$	0.0389 ± 0.0068	μM

3.6 Supplementary Information for Effective Biophysical Modeling of Cell Free Transcription and Translation Processes

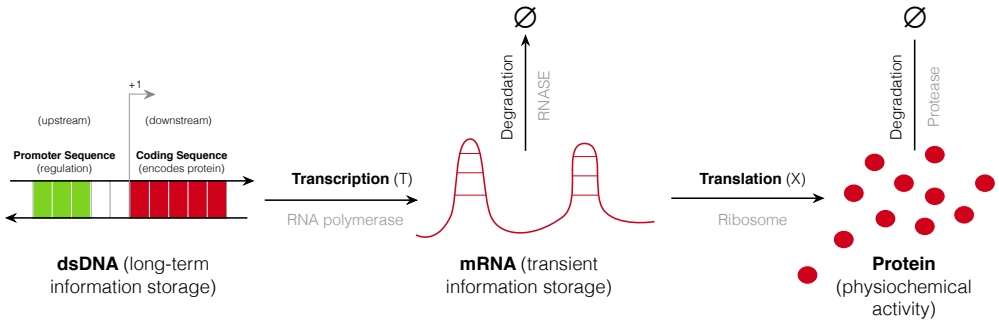


Figure 3.4: Schematic of transcriptional (TX) and translation (TL) processes.

Effective expressions for transcription (and translation) kinetics

To develop expressions for $r_{X,i}$ (or $r_{L,i}$) let us develop a mental model of the elementary steps occurring in transcription (Fig. 3.4). Our mental model for transcription, based upon the earlier work by McClure [136] and later Bailey [115], consists of a four step elementary reaction scheme:

$$\mathcal{G}_j + R_X \rightleftharpoons \left(\mathcal{G}_j : \mathrm{R_X}\right)_C \tag{3.15}$$

$$\left(\mathcal{G}_j : \mathrm{R_X}\right)_C \longrightarrow (\mathcal{G}_j : \mathrm{R_X})_O \tag{3.16}$$

$$\left(\mathcal{G}_j : \mathrm{R_X}\right)_O \longrightarrow R_X + \mathcal{G}_j \tag{3.17}$$

$$(\mathcal{G}_j : \mathrm{R_X})_O \longrightarrow m_j + R_X + \mathcal{G}_j \tag{3.18}$$

where $\mathcal{G}_j$, R_X denote the gene and *free* RNA polymerase (RNAP) concentration, and $\left(\mathcal{G}_j : \mathrm{R_X}\right)_O$, $\left(\mathcal{G}_j : \mathrm{R_X}\right)_C$ denote the open and closed complex concentrations, respectively. Let the kinetic rate of transcription be directly proportional to the

concentration of the open complex: [backgroundcolor=lgray]

$$r_{X,j} = k_{E,j}^{X} \left(\mathcal{G}_j : \mathrm{R_X} \right)_O$$

where $k_{E,j}^{X}$ is the elongation rate constant for gene j. The key idea behind this derivation is that the RNAP (or Ribosome) acts as an enzyme. Thus, we might expect that we could use a strategy similar to enzyme kinetics to derive an expression for $r_{X,j}$ (and $r_{L,j}$). The material balances around the closed and open complex for gene j are given by:

$$\frac{d}{dt}\left(\mathcal{G}_j : \mathrm{R_X} \right)_C = k_+ \left(\mathcal{G}_j \right) (\mathrm{R_X}) - k_- \left(\mathcal{G}_j : \mathrm{R_X} \right)_C - k_I \left(\mathcal{G}_j : \mathrm{R_X} \right)_C \tag{3.19}$$

$$\frac{d}{dt}\left(\mathcal{G}_j : \mathrm{R_X} \right)_O = k_I \left(\mathcal{G}_j : \mathrm{R_X} \right)_C - k_A \left(\mathcal{G}_j : \mathrm{R_X} \right)_O - k_{E,j}^{X} \left(\mathcal{G}_j : \mathrm{R_X} \right)_O \tag{3.20}$$

where k_+ ($\mathrm{conc}^{-1}\ \mathrm{t}^{-1}$) and k_- (t^{-1}) denote the on/off rate constant for RNAP at the promoter for gene j, k_I (t^{-1}) denotes the rate constant governing open complex formation and k_A (t^{-1}) denotes the rate constant governing abortive initiation. The total abundance of RNAP, denoted as $R_{X,T}$ is governed by:

$$R_{X,T} = R_X + \left(\mathcal{G}_j : \mathrm{R_X} \right)_C + \left(\mathcal{G}_j : \mathrm{R_X} \right)_O \tag{3.21}$$

At steady state, the abundance of the closed and open complexes can be estimated from the balance equations (where we have neglected the subscript j for simplicity):

$$\left(\mathcal{G}_j : \mathrm{R_X} \right)_C \simeq \left(\frac{k_+}{k_- + k_I} \right) \left(\mathcal{G}_j \right) (\mathrm{R_X}) \tag{3.22}$$

$$\left(\mathcal{G}_j : \mathrm{R_X} \right)_O \simeq \left(\frac{k_I}{k_A + k_E^X} \right) \left(\mathcal{G}_j : \mathrm{R_X} \right)_C \tag{3.23}$$

The ratio of parameters in the open and closed complex expressions have special significance which is apparent from looking at their units. For example, the ratio:

$$K_{X,j}^{-1} \equiv \left(\frac{k_+}{k_- + k_I} \right) \tag{3.24}$$

is a saturation constant for gene j with units of concentration, while:

$$\tau_{X,j}^{-1} \equiv \left(\frac{k_I}{k_A + k_E^X} \right) \tag{3.25}$$

is a time constant for gene j comparing the initiation, abortive initiation and elongation constants. We can relate the open complex to the concentration of gene j and *free* RNAP concentration by eliminating the closed complex concentration from the steady state expressions:

$$\left(\mathcal{G}_j : \mathrm{R_X} \right)_O \simeq (K_{X,j}^{-1})(\tau_{X,j}^{-1}) \left(\mathcal{G}_j \right) (\mathrm{R_X}) \tag{3.26}$$

To estimate the *free* RNAP concentration we can use the total RNAP balance, where we have substituted expressions for the open and closed complex concentrations:

$$R_{X,T} = R_X + (K_{X,j}^{-1}) \left(\mathcal{G}_j \right) (\mathrm{R_X}) + (K_{X,j}^{-1})(\tau_{X,j}^{-1}) \left(\mathcal{G}_j \right) (\mathrm{R_X}) \tag{3.27}$$

Starting with Eqn (3.27), solving for *free* RNAP concentration R_X gives:

$$R_X = \frac{R_{X,T} \left(\tau_{X,j} K_{X,j} \right)}{\tau_{X,j} K_{X,j} + \left(\tau_{X,j} + 1 \right) \mathcal{G}_j} \tag{3.28}$$

Now that we have R_X we can get the open complex concentration in terms of total RNAP:

$$\left(\mathcal{G}_j : \mathrm{R_X} \right)_O \simeq \frac{R_{X,T} \mathcal{G}_j}{\tau_{X,j} K_{X,j} + \left(\tau_{X,j} + 1 \right) \mathcal{G}_j} \tag{3.29}$$

Lastly, the kinetic rate of transcription is proportional to the open complex concentration which can now be substituted to give:

$$r_{X,j} = k_{E,j}^X R_{X,T} \left(\frac{\mathcal{G}_j}{\tau_{X,j} K_{X,j} + \left(\tau_{X,j} + 1 \right) \mathcal{G}_j} \right) \tag{3.30}$$

In an identical procedure, we can also formulate a model of the translation rate:

$$r_{L,j} = k_{E,j}^L R_{L,T} \left(\frac{m_j}{\tau_{L,j} K_{L,j} + \left(\tau_{L,j} + 1 \right) m_j} \right) \tag{3.31}$$

where m_j denotes the concentration of mRNA j.

Which is limiting, elongation or initiation?

Ultimately, this question depends upon the gene of interest. However, we can see some interesting properties of $r_{X,j}$ by considering limiting cases for the value of the time constant $\tau_{X,j}$. Assume the rate constant for abortive initiation k_A is small compared to both k_I and $k^X_{E,j}$:

$$\tau_{X,j} \simeq \frac{k^X_{E,j}}{k_I} \tag{3.32}$$

When $\tau_{X,j} \gg 1$ (initiation limited) the kinetic transcription rate becomes:

$$r_{X,j} = \frac{k^X_{E,j} R_{X,T}}{\tau_{X,j}} \left(\frac{\mathcal{G}_j}{K_{X,j} + \mathcal{G}_j} \right) \tag{3.33}$$

while $\tau_{X,j} \ll 1$ (elongation limited) gives:

$$r_{X,j} = k^X_{E,j} R_{X,T} \left(\frac{\mathcal{G}_j}{K_{X,j}\tau_{X,j} + \mathcal{G}_j} \right) \tag{3.34}$$

How do we get values for k_+, k_-, k_I, $k^X_{E,j}$ and k_A?

Generally speaking, except for $k^X_{E,j}$ which we can estimate from first principles, estimating the value of k_+, k_-, k_I and k_A is difficult (especially *in-vivo*). Thus, let's start with $k^X_{E,j}$; the elongation rate constant is proportional to the elongation rate of the polymerase e_X (units of nt s^{-1}) multiplied by the length (nt) of the coding region of gene j, or $\mathcal{L}_j$ (the length of DNA the RNAP has to read). However, typically we formulate $k^X_{E,j}$ in a slightly different way; first, we compute an average or characteristic elongation rate constant $\left\langle k^X_E \right\rangle$, and then correct this characteristic value by the actual length of gene j:

$$k^X_{E,j} = \left\langle k^X_E \right\rangle \left(\frac{\mathcal{L}}{\mathcal{L}_j} \right) \tag{3.35}$$

where:

$$\left\langle k_E^X \right\rangle = e_X \mathcal{L}^{-1} \tag{3.36}$$

and $\mathcal{L}$ denotes some characteristic length, e.g., the average length of genes in *E.coli*. For the other parameters, we must estimate them from experimental data.

McClure performed a series of *in vitro* experiments to estimate k_I in transcription and produced a constraint governing permissible values for the remaining transcriptional parameters [136]. In particular, McClure used an abortive initiation assay in which the production of mRNA never completed. Instead, transcription always aborted leaving a stable open complex that could be directly measured. From these measurements, and a mental transcriptional model very similar to ours, McClure developed the expression:

$$\tau_{obs} = \frac{1}{k_I} + \frac{1}{R_{X,T}} \left(\frac{k_- + k_I}{k_+ k_I} \right) \tag{3.37}$$

where τ_{obs} is the time required to fully form the open complex (measured) and $R_{X,T}$ denotes the total concentration of RNAP. A value for k_I, and a relationship between the other parameters, can be obtained from the intercept and slope of a $R_{X,T}^{-1}$ versus τ_{obs} plot for a particular promoter of interest.

3.7 Acknowledgement

The work described was also supported by the Center on the Physics of Cancer Metabolism through Award Number 1U54CA210184-01 from the National Cancer Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

CHAPTER 4

DEVELOPING A GLUCONATE-RESPONSIVE *E. COLI* TRANSCRIPTIONAL ELEMENT IN AN IN VITRO SYSTEM

4.1 Introduction

A number of effective metabolic engineering strategies have utilized the advances in promoter engineering to dynamically regulate the expression of pathway enzymes to optimize product titers [112, 120, 55, 121]. A core component of such a scheme is an allosteric transcription factor (aTF)– a protein which undergoes a conformational change upon binding to its specific ligand, significantly changing its DNA binding ability. Ligand binding leads to the release of the aTF from the operator, thereby allowing the RNA polymerase holoenzyme to effectively bind to the promoter and increase the transcription rate. More recently, due to advances in the cell free protein synthesis (CFPS) platform, aTFs have been increasingly leveraged for point-of-care biosensing purposes. CFPS is not only attractive for constructing such biosensors because of its fast design-build-test cycle, but it is also a promising platform for implementing them because of its portability, ease of fine-tuning the reaction components, and tolerance in cytotoxic environments, enabling its point-of-care use. In these applications, the aTFs are used to 'sense' a specific molecule (ligand) by altering gene expression of a downstream reporter protein, the signal of which would be proportional to the amount of ligand present. Such biosensor systems have recently been used for detecting ions, biomarkers, and antibiotics [201, 239]. Taken together, these aTfs can be effectively used to strategically alter gene expression for a variety of applications.

We are interested in implementing an aTF that can respond to glucose or its derivatives in a genetic circuit in CFPS to create an input-responsive, on-demand, protein synthesis system. Such a system could also be used in medical applications, in particular to manage Type I and II diabetes. By 2045 the total cases of diabetes is expected to rise to 629 million[1], and there is a pressing demand to develop effective strategies to manage this global disease. The ability to sense glucose concentration in the bloodstream and direct the synthesis of either insulin or glucagon hormone (or a combination of both) via the use of genetic circuits is very attractive, and given the advanced state of cell free systems, it is an achievable feat. Toward this aim, it is important to first construct a transcriptional unit that can effectively respond to varying levels of glucose or its derivatives that can act as its proxy.

We plan to approach this problem by making use of aTFs. The aTF protein has two main domains: a DNA binding region in the N-terminus and a ligand binding region in the C-terminus. Numerous studies on the bacterial regulators such as TetR, GntR and ArgR have shown that these protein families have a relatively conserved sequence for the N-terminal region [37, 43, 209, 32]. These proteins bind to specific operator regions on the DNA via a helix-turn-helix (HTH) motif. On the other hand, the C-terminal sequence varies for each member in the subfamily, giving rise to different ligand binding specificities. Studies on the gluconate operon of *B. subtilis* and *E. coli* have shown that two derivatives of glucose, D-glucono-1,5-lactone, and more so D-gluconate, act as antirepressors, relieving the repression by GntR protein [142, 57, 169, 178, 226, 255, 195]. For our application, we therefore focused on the GntR transcription factor family.

[1]Diabetes facts and figures. International Diabetes Federation. Available from: idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html

In this work, we aimed to first construct a promoter (incorporating the GntR aTF) that can be efficiently used in an *E. coli*-based CFPS. Toward this aim, we first cloned the GntR operator sequence to a bacteriophage lambda promoter (P70) developed by Noireaux and coworkers [191]. The P70 promoter construct, along with its 5′ untranslated region, was chosen because it is optimized to maximize expression in the *E. coli* myTXTL CFPS system [60]. Next, we added a reporter protein gene downstream of this modified promoter to easily characterize the promoter dynamics. The CFPS reactions were carried out either in the absence or presence of the GntR repressor gene. We characterized the repression and de-repression characteristics (using D-gluconate) in two different *in vitro* systems: an extract-based *E. coli* system, myTXTL, and a reconstituted system derived from PURE [189]. Taken together, our results show that the GntR protein effectively represses the expression of the reporter protein, and that the addition of D-gluconate relieves the repression.

4.2 Materials and Methods

4.2.1 Construction of a modified bacterial promoter regulated by a gluconate-responsive repressor

The first step in constructing the promoter responsive to D-gluconate levels was to include the operator sequence for the repressor, GntR (*B. subtilis* or *E. coli*), in the promoter region. The sequences of the operator and the repressor were obtained from the literature [142, 57, 169, 178, 226, 255, 195].

E. coli promoter construction

The 16bp *E. coli* GntR operator sequence (ATGTTACCCGTATCAT) was added between the -35 and -10 region of mutated bacteriophage lambda promoter (P70) developed by Noireaux and coworkers [191] for efficient expression in the myTXTL cell free system. The -10 binding site of the RpoD holoenzyme was altered from GATAAT in the original promoter to TATCAT in the modified P70 promoter. The gene sequence of the reporter protein, deGFP (or Venus), was added downstream of this modified promoter. On the other hand, the GntR repressor gene was cloned downstream of the original (P70) promoter. The terminator and 5′ untranslated sequences for the modified P70 DNA were identical to those of the P70 DNA. The full constructs were ordered as linear DNA fragments, with 150-200 bp flanker sequences on both ends, from Twist Bioscience and Integrated DNA Technologies.

B. subtilis promoter construction

The 32 bp *B. subtilis* GntR operator sequence (CGTTATCATACTTGTATA-CAAGTATACTCCTT) was added after the -10 region of the bacteriophage lambda promoter developed by Noireaux and coworkers [191]. The gene sequence of the reporter protein, deGFP, was added downstream of this modified promoter. The full constructs were ordered as linear DNA fragments, with 150-200 bp flanker sequences on both ends, from Integrated DNA Technologies.

4.2.2 Cell free synthesis in myTXTL extract

The cell free protein synthesis (CFPS) reactions were carried out using the myTXTL Sigma 70 Master Mix (Arbor Biosciences) at 29 °C in 384 well plates (Thermofisher NUNC, flat bottom) in the Varioskan Lux plate reader. The working volume of all the reactions was approximately 15 µL, composed of the Sigma 70 Master Mix (9 /*mu*L), GamS nuclease inhibitor (5 μM), and the linear DNA fragments: modifiedP70-deGFP (10 nM) or P70-deGFP (10nM), GntR (5nM). For the control reactions without GntR, equal volume of nuclease-free water was used instead. For the de-repression experiments, 0.5 μL D-gluconate was added to the extract at a final concentration of 10 mM.

4.2.3 Cell free synthesis in PURE extract

These CFPS reactions were carried out in the reconstituted PURExpress system (New England Biolabs) at 37 °C in 384 well plates (Thermofisher NUNC, flat bottom) in the Varioskan Lux plate reader. The working volume of all the reactions was approximately 15 µL, composed of solution A and B mixed according to the manufacturer's instructions, RNAse inhibitor (20 Units per reaction), NEB *E. coli* RNA polymerase holoenzyme (1-2 μL), and the linear DNA fragments added at different concentrations. For the control reactions without GntR, equal volume of nuclease-free water was used instead. For the de-repression experiments, 0.5 μL D-gluconate was added to the extract at a final concentration of 10 mM.

4.2.4 Measurement of protein levels

The Green fluorescent protein (deGFP) [191] fluorescence was measured using the Varioskan Lux plate reader at 488 nm (excitation) and 535 nm (emission). The Venus fluorescent protein was measured in the same instrument at 513 nM (excitation) and 531 nM (emission).

4.2.5 Measurement of D-gluconate levels in the cell free extract

myTXTL CFPS reactions were carried out with or without glucose oxidase/glucose for the required time intervals in 1.5 mL Eppendorf tubes at 29 °C. The CFPS reaction mix either contained 2 μL of 0.025 U/μL glucose oxidase or the same amount of water. In addition, 1 μL glucose was added to a final concentration of 166 mg/dL. For the control case, 1 μL water was added instead of glucose.

D-gluconate levels were quantified by Liquid Chromatography/Mass Spectrometry (LCMS) using our published derivatization protocol [235]. Briefly, end-point CFPS reactions were quenched with 100% ice-cold ethanol in a 1:1 volumetric ratio. Ethanol precipitated samples were centrifuged at 12000×g for 15 minutes at 4 °C. The supernatants were collected and tagged with C-12 aniline, using EDC (N-(3-dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride) as a catalyst. After 2 hours of gentle shaking, 1.5 μL of TEA (triethylamine) was added, the mixture centrifuged at 13500×g for 3 minutes and the supernatant analyzed using the LCMS system. Meanwhile, D-gluconate standards at different concentrations were similarly tagged with aniline and quantified. A standard curve was used to determine absolute concentrations of D-gluconate.

4.3 Results and Discussion

Quantification of D-gluconate concentration

D-gluconate acts a dynamic second messenger for D-glucose concentration in myTXTL cell free extracts (Fig. 4.1). We used the gluconate-responsive regulator GntR as the key transcription element for the glucose sensor circuit. Thus, information about the background D-gluconate concentration, and its temporal evolution in response to changes in glucose concentration, is critical to the glucose sensing moiety of the proposed circuit. Toward this question, we measured the D-gluconate concentration before and after the addition of exogenous D-glucose, in the presence and absence of Glucose Oxidase (EC 1.1.3.4) in myTXTL cell free extracts. We added approximately 166 mg/dl of D-glucose (the higher end of the normal blood glucose range of a diabetic patient after a meal) at t = 1.25h and monitored the D-gluconate concentration by LCMS using our previous derivatization protocol [235]. Following the addition of exogenous D-glucose, in the presence of glucose oxidase, the D-gluconate concentration peaked sharply to approximately 590 μM within five minutes of the addition of glucose, and then slowly decayed to the initial steady state value over two hours (Fig. 4.1, solid lines). A similar trend was observed in the absence of glucose oxidase, albeit with a lower peak concentration (Fig. 4.1, dashed lines). Taken together, this study suggested there was native glucose oxidase/dehydrogenase activity in the myTXTL extract that can convert D-glucose to D-gluconate, the D-gluconate signal abundance could be improved with the external addition of glucose oxidase, and D-gluconate was consumed by downstream metabolic processes leading to a transient signal.

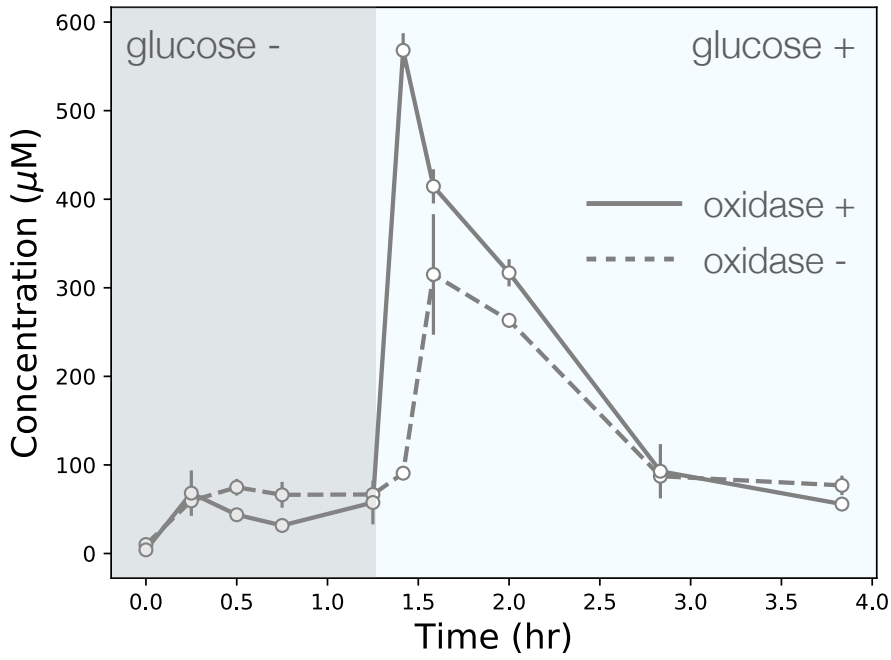


Figure 4.1: D-gluconate quantification. Glucose pulse and gluconate consumption experiments in myTXTL cell free extract. Glucose with a concentration approximating that of a typical diabetic patient (160 mg/dl) was added to the cell free reaction mixture at t = 1.25 hours. D-gluconate concentration, a downstream degradation product, was measured by LCMS with and without the addition of Glucose oxidase to the cell free mixture. Points denote the mean while the error bars denote one standard deviation

Repression and de-repression characteristics of GntR in the myTXTL extract

We first tested the performance of linear DNA constructs in the myTXTL extract using a P70-deGFP construct (10nM) in the presence of GamS nuclease inhibitor (5 μM) over 16 hours (Fig. 4.2). The deGFP levels were comparable to those using a plasmid version of P70-deGFP (5nM) in our previous study [4]. This established the feasibility of the use of linear DNA for our experiments. Next, we characterized repression by GntR in the myTXTL extract (Fig. 4.3). We expressed two linear DNA fragments: *E. coli* GntR (2.5nM) and deGFP placed downstream of the modified P70 promoter (10nM). For the control case, the GntR DNA was excluded from the reaction mixture. While the deGFP lev-

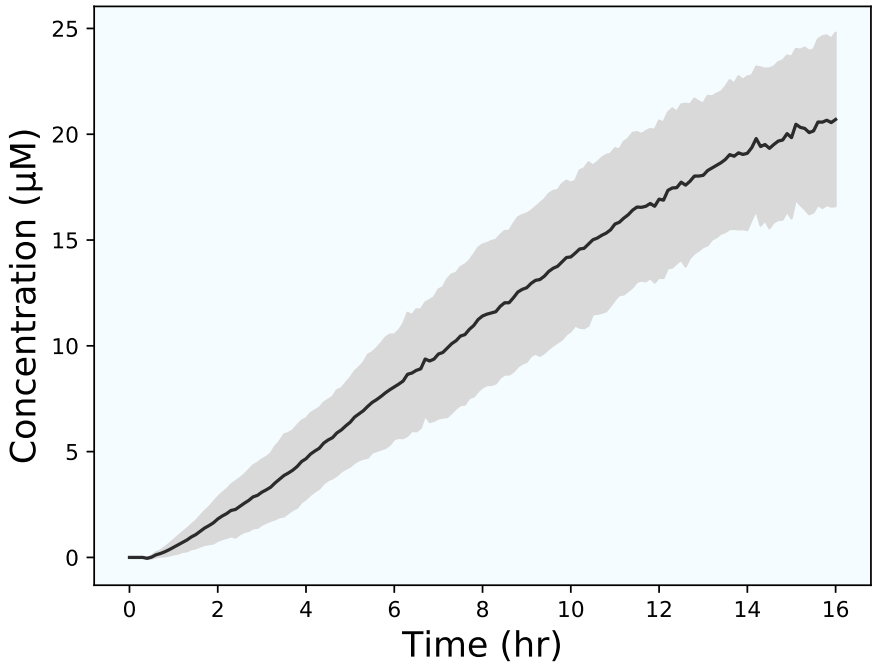


Figure 4.2: deGFP yield using 10 nM P70-deGFP linear DNA in myTXTL extract. The reaction was supplemented with 5μM GamS nuclease inhibitor. Shaded region denotes one standard deviation.

els in the control case steadily increased, these levels reached a steady plateau in approximately 4 hours in the +GntR reaction. This showed that GntR was indeed repressing the expression of deGFP. In order to test the de-repression characteristic of GntR, we added 10mM of D-gluconate (approximately 9-fold the value of its K_D established in literature [38]) after deGFP levels plateaued (4 hours). The deGFP levels increased compared to the case where gluconate was not added; this increase was especially apparent in the peak that resulted after an hour for the +gluconate case (Fig. 4.4). However, the overall increase in deGFP levels was too low compared to the −GntR control, highlighting that full de-repression did not take place. The most probable cause for the poor de-repression is the metabolic consumption of D-gluconate by the myTXTL extract. The addition of gluconate to the extract also led to a slightly decreased

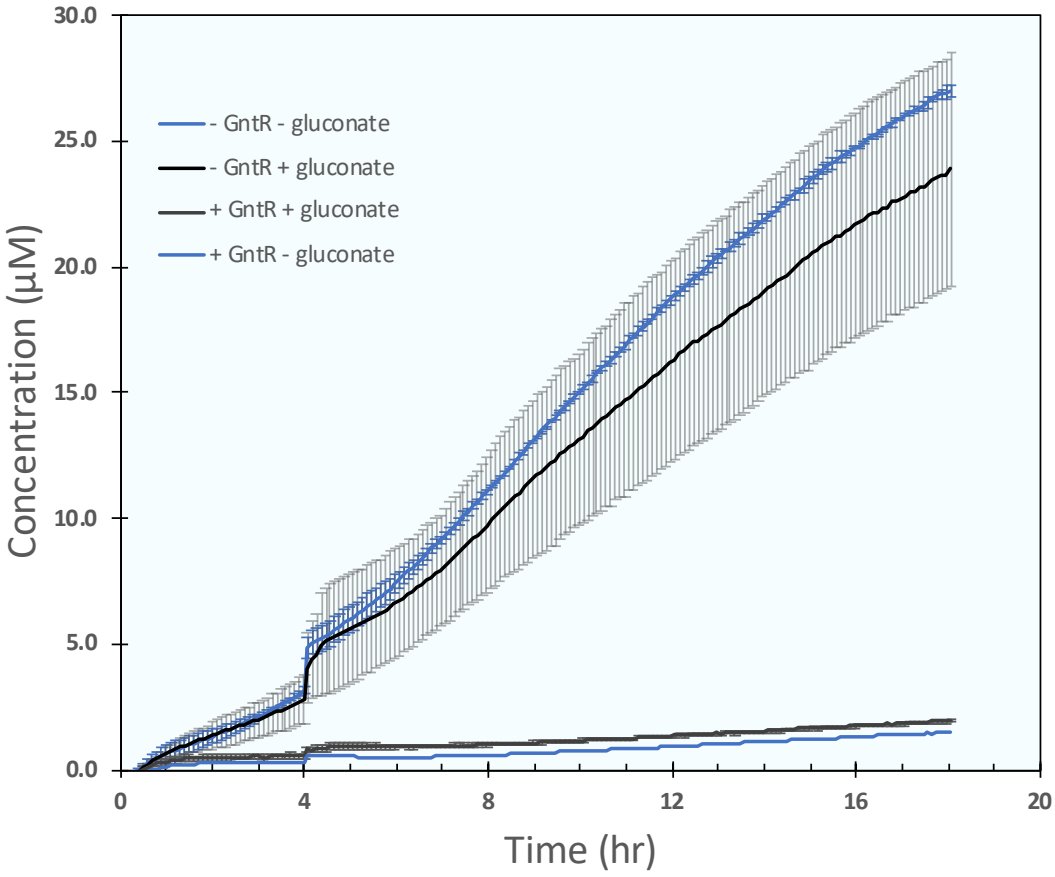


Figure 4.3: Repression of deGFP expression by *E. coli* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (2.5 nM) Error bars denote one standard deviation.

final yield of deGFP (compared to the −GntR−gluconate case), although this decrease was not significant. Additionally, although the -10 region of the modified P70 promoter (TATCAT) was altered from the original P70 promoter (GATAAT), it did not negatively impact the transcription, as can be seen from (Fig. 4.2 and 4.3).

We also tested a GntR repressor derived from *B. subtilis* in the myTXTL extract (Fig. 4.5). In this version, the promoter construct was different from the previous case (*E. coli*), where the operator was placed between the -35 and -10

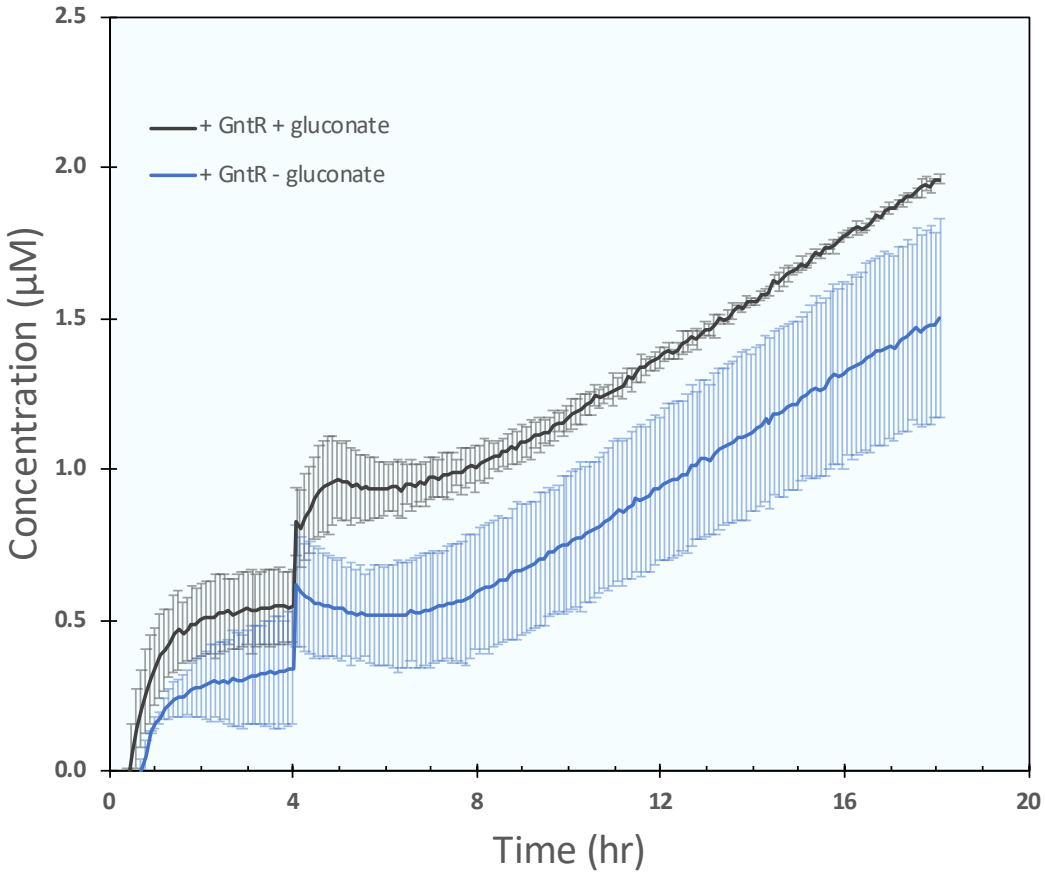


Figure 4.4: Repression of deGFP expression by *E. coli* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (2.5 nM) Error bars denote one standard deviation.

region of the promoter. Because the operator sequence for the *B. subtilis* GntR was 32bp in length, we placed it after the -10 region. However, in this case, we observed very low levels of deGFP expression. We also conducted similar derepression experiments using D-gluconate; however, the deGFP levels were too low (they were in the noise region) to confidently come to a conclusion. Possible reasons for the low expression of deGFP are the poor binding of the RNA polymerase to the promoter or an unfavorable 5′ untranslated region including the ribosome binding site. We plan to conduct future experiments with the *B. subtilis* GntR in a different promoter construct.

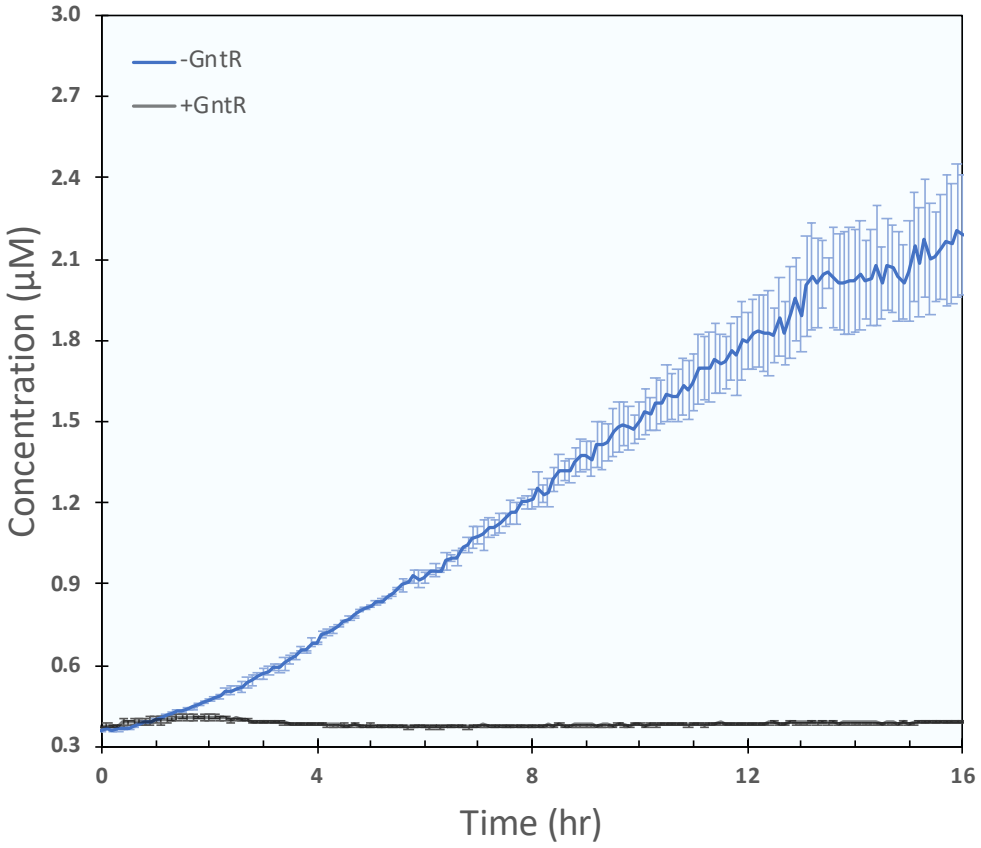


Figure 4.5: Repression of deGFP expression by *B. subtilis* GntR in myTXTL extract. DNA fragments used: ModifiedP70-deGFP (10 nM), GntR (5 nM) Error bars denote one standard deviation.

final yield of deGFP (compared to the – GntR –gluconate case), although this decrease was not significant. Additionally, although the -10 region of the modified P70 promoter (TATCAT) was altered from the original P70 promoter (GATAAT), it did not negatively impact the transcription, as can be seen from (Fig. 4.2 and 4.3).

Lastly, we tested if the *B. subtilis* GntR protein showed any affinity for the operator sequence for the *E. coli* version. For this experiment, we used equimolar DNA concentrations (8nM) for the GntR gene and the promoter. The control case lacked the GntR gene. Our results showed no cross-talk between the two

versions. In addition, this data also showed that the expression of an additional protein did not affect the expression of deGFP, highlighting the presence of adequate amount of transcription and translation resources in the extract. Taken together, our studies of GntR in the myTXTL extract showed effective repression behavior.

Repression and de-repression characteristics of GntR in the reconstituted PURE system

The GntR repression experiments (using *E. coli*) parts were carried out next in a reconstituted cell free system lacking metabolic capability to avoid the unintended consumption of D-gluconate (Fig. 4.6). All results for this system were expressed in relative fluorescence units (RFU). In addition to deGFP, another fluorescent reporter, Venus, was also tested (using the same promoter construction) because it was previously shown that Venus is a much brighter reporter protein in this system, and it had similar time at maximum growth values to different mutated GFP variants [116]. In agreement with previous literature, the deGFP reporter fluorescence was low in our experiments (data not shown). As a result, subsequent experiments were carried out using Venus. Different concentrations of the Venus reporter construct with or without the presence of GntR were first tested (Fig. 4.6). The use of 5nM Venus and 12nM GntR DNA resulted in better repression (Fig. 4.6 B), and this configuration was thus used in a following experiment to determine the de-repression by gluconate. The de-repression experiment was carried out using the same amount of D-gluconate (10mM) as in the myTXTL case; however, in this case, it was added to the reaction mixture at time 0, assuming that it would not be consumed by the cell

free system (Fig. 4.7). The results of this experiment showed that D-gluconate acted as an antirepressor, increasing the expression of Venus. While the levels of Venus did not reach those of the -GntR case, it was still appreciably higher than the +GntR −gluconate case. Taken together, GntR was shown to effectively repress expression in the reconstituted cell free extract.

4.4 Conclusion

This study investigated the repression and de-repression performance of the GntR aTF derived from two different prokaryotes, *E. coli* and *B. subtilis*. The GntR aTF was chosen for our study because it was previously shown to be responsive to D-gluconate, a close derivative and proxy of D-glucose. The promoter regions of the two variants were constructed based on the transcription by the *E. coli* RNA polymerase (bound to sigma 70), and they included the operator sequence for the respective GntR protein version. Gene expression using the *B. subtilis* version of the construct was low, putatively due to the reduced binding affinity of the RNA polymerase to the promoter region or the ribosome to the ribosome binding site. The *E. coli* version of the construct showed promising repression and de-repression results, especially in the reconstituted extract, which did not have metabolic side reactions that consumed D-gluconate. Taken together we have constructed a promising transcriptional unit that can ”sense” glucose via its proxy, D-gluconate.

There are several future directions for this work that we plan to pursue. First, in order to fully characterize repression and de-repression, we plan to perform (i) electrophoretic mobility shift assays (EMSA) using the promoter DNA and

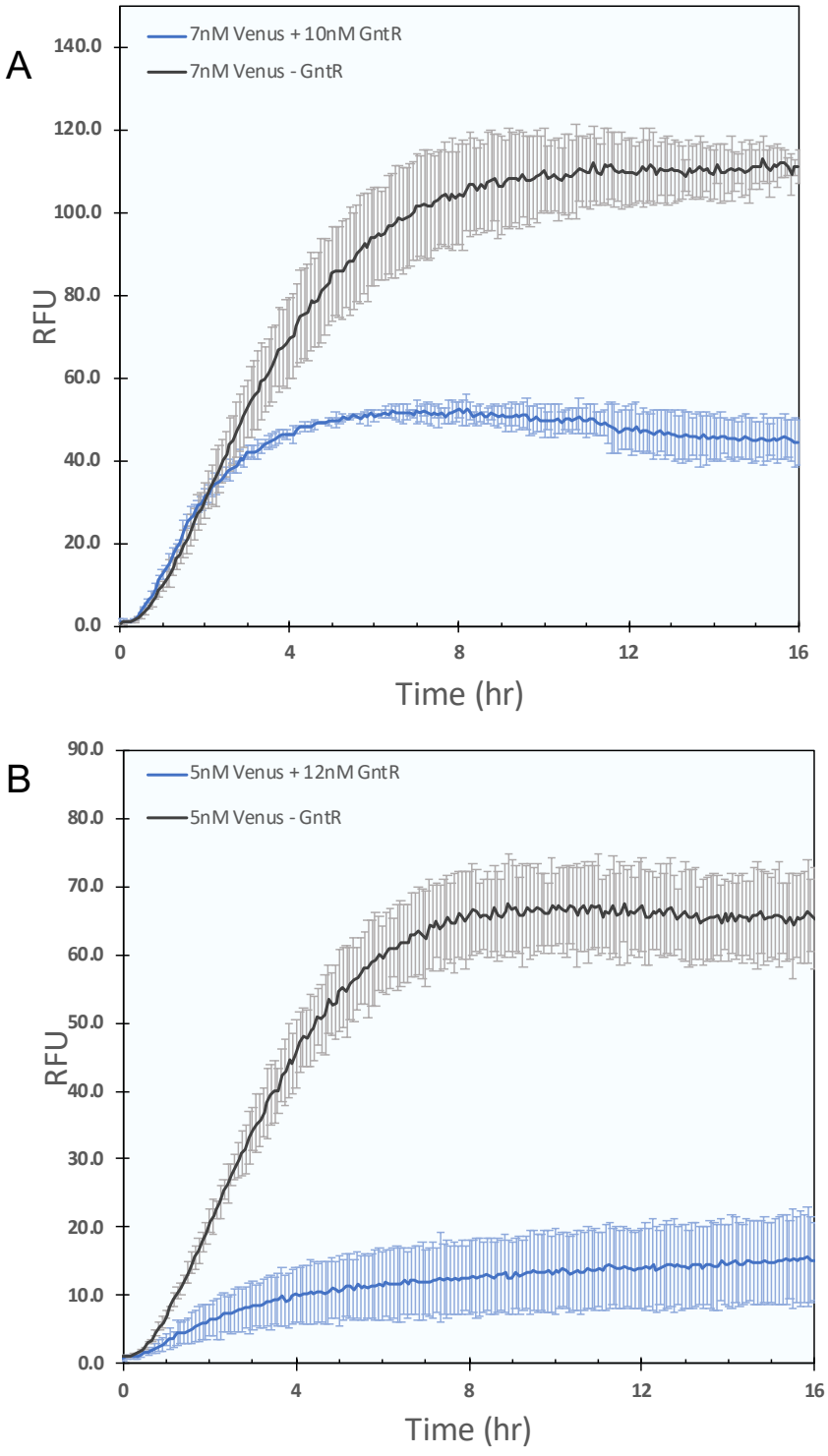


Figure 4.6: Repression of Venus expression by *E. coli* GntR in PURE system. Error bars denote one standard deviation.

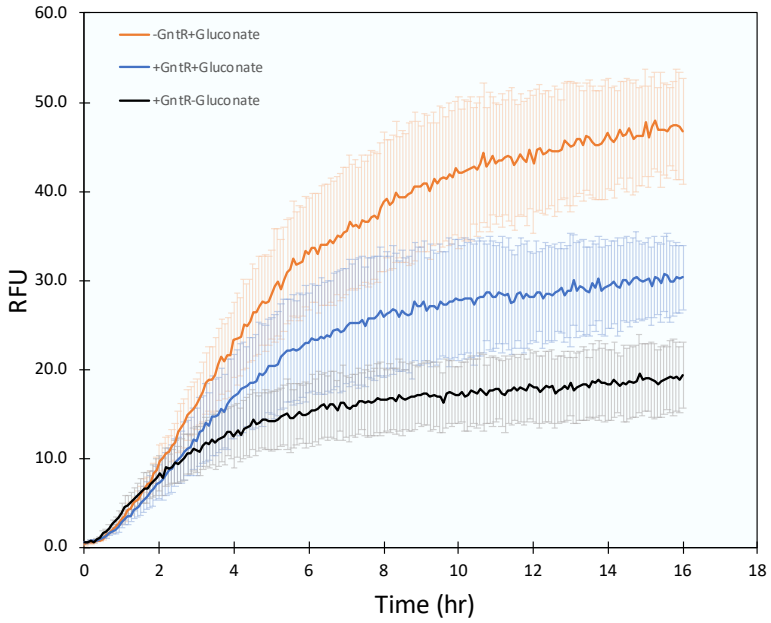


Figure 4.7: De-repression of Venus expression by *E. coli* D-gluconate in PURE system. 10mM D-gluconate was added to the reaction mixture at time 0. Error bars denote standard error.

the GntR protein in varying concentrations, (ii) end point measurements of the reporter fluorescence in the reconstituted system using different levels of promoter DNA and GntR protein. We also plan to use a different GntR constuct, derived from *P. aeruginosa* [38] in the EMSA and end point measurements. This construct was shown to be partly responsive to D-glucose, eliminating the need for a proxy, D-gluconate. Second, we plan to determine Venus reporter levels in concentration units to determine actual protein yield. This would be done using a calibration curve of the purified Venus protein. Third, we plan to develop a mathematical model for this system to effectively characterize transcription and translation behavior; such a model would also help guide our future experiments based on this system. Lastly, we plan to further engineer the GntR protein (and the promoter region) to enable more efficient repression and de-repression properties.

CHAPTER 5

CONCLUSION AND FUTURE DIRECTIONS

In this work, we developed a rigorous framework toward extending the capability of cell free protein synthesis (CFPS). First, we developed an effective biophysical modeling approach to simulate transcription and translation processes involving genetic circuits in CFPS. Global sensitivity analysis was also embedded in the model to study the influence of individual model parameters on the expression dynamics. This model can be easily extended to include a more intricate circuit architecture. Second, we developed a high-throughput chromatography method for the absolute quantification of metabolites involved in the *E. coli* central carbon pathway. This method can be used to experimentally validate the computational flux balance approaches developed in our group in the study of metabolism, which is increasingly important for Synthetic Biology applications involving modern extract-based cell free systems. Metabolism is tied closely to energy regeneration in these systems, and its optimization would not only enable higher protein yields but also improve the longevity of the CFPS reactions. Third, we constructed a transcriptional unit based on a gluconate-responsive allosteric transcription factor, GntR, that can be utilized in glucose-sensing applications. Our preliminary results of this transcriptional unit showed encouraging results. We aim to incorporate this unit in a more intricate genetic circuit to fully utilize its glucose-sensing potential, for example, to direct downstream synthesis proteins such as insulin and glucagon for the management of Type I and II diabetes. Moreover, the modeling approaches developed in our group can be utilized for the design and implementation of this genetic circuit. Taken together, this body of work has laid an important foundation toward the development of a prototype of an *in vitro* model for pancreas.

5.1 Future Directions

5.1.1 Error-Prone PCR-Based Mutagenesis of GntR repressor protein and in vivo screening

In this work, we showed that D-gluconate acted as an antirepressor for GntR; however, the expression was not fully restored to fully de-repressed levels. Thus, the de-repression property can be further improved. In order to increase the antagonizing effect of D-gluconate on the GntR repressor binding, we propose to synthesize a library of GntR repressor mutants using error-prone PCR and screen for the mutants with high affinity for the effector. The GntR protein and its subfamily has been studied and the DNA as well as effector binding motifs have been predicted by several research groups [79, 179, 1]. In particular, Fujita and coworkers [254] determined that mutations in the C-terminal region of GntR affected its ability to bind to D-gluconate. A previous study from the same group showed that N-terminal residues were responsible for operator binding properties [256]. On the basis of these details, we plan to generate a library of GntR protein with random mutations in the C-terminal region via error-prone PCR. The selection procedure (*in vivo*), with the goal of discovering GntR mutants that (i) do not have operator binding strength compromised by the mutations, and (ii) are effectively antagonized by D-gluconate, will be as follows. *E. coli* strain devoid of genes encoding for the enzymes involved in D-gluconate consumption or production (glucose dehydrogenase [EC 1.1.5.2], gluconate 2-dehydrogenase [EC 1.1.1.215] and gluconokinase [EC 2.7.1.12]) will be used for the experiments. The cells will be doubly transformed with compatible plasmids: first with a plasmid carrying the GntR mutant under a constitutive bacte-

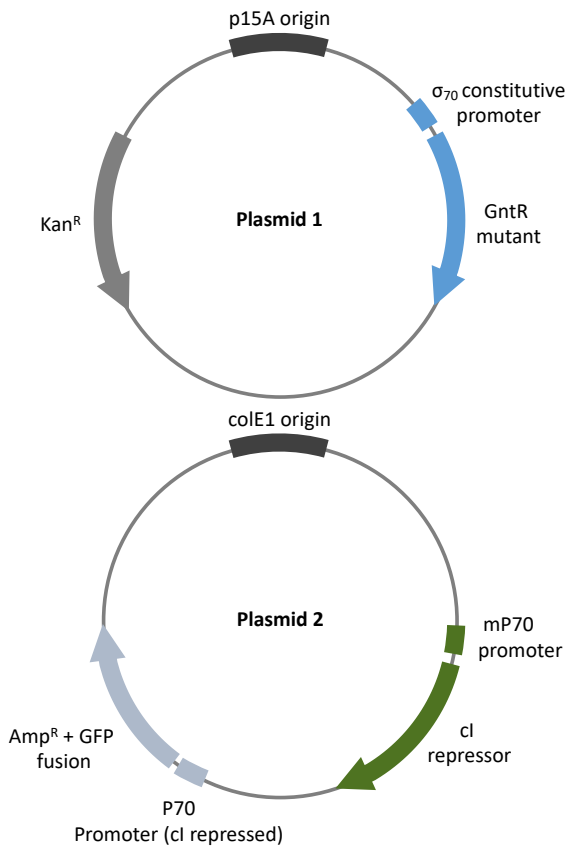


Figure 5.1: Map of plasmids used for the screening and selection procedure. *E. coli* cells used for the procedure will be doubly transformed with these plasmids. mP70 is the modified P70 promoter with added GntR operator site.

rial promoter and a Kanamycin resistance gene (Kan^R); second with a plasmid carrying two genes: (i) fluorescent reporter/antibiotic resistance fusion gene, for example Ampicillin resistance gene (Amp^R) fused with GFP under a P70 bacterial promoter, and (ii) a cI repressor gene under the gluconate-responsive promoter, mP70 (Fig. 5.1). The doubly transformed cells will initially be grown in agar plates (without D-gluconate) carrying the two antibiotics. Survival of the cells depends on the repression of cI protein (that represses Amp^R expression), which in turn relies on the repression efficiency of GntR. Thus, this initial selection will yield only colonies that carry GntR mutants that effectively bind to the operator. The colonies that have the brightest reporter fluorescence will

be picked for a screening round, wherein they will be grown in liquid media in the presence of D-gluconate and only Kanamycin antibiotic, and analyzed by flow cytometry with Fluorescence Activated Cell Sorting (FACS). The cells will be binned and sorted into different categories according to reporter fluorescence measurement thresholds. Low fluorescence will correlate with high D-gluconate binding affinity of GntR protein. The lowest fluorescing cells will be determined following a second round of FACS, and its plasmids will be sequenced to determine GntR sequence. This procedure will be iterated until the most efficient GntR protein candidate is discovered.

5.1.2 Develop and characterize a glucose-responsive genetic switch in an *E. coli* cell free system

We want to build on our current work on by incorporating the glucose-responsive transcriptional unit in a genetic switch circuit in an *E. coli* CFPS (Fig. 5.2). Because our unit depends on D-gluconate, we want to be able to first convert glucose to D-gluconate for it to act as a proxy for glucose. This will be done by adding glucose oxidase enzyme (EC 1.1.3.4) to the extract to convert glucose to D-gluconate. Alternatively, glucose can also be converted using glucose dehydrogenase (EC 1.1.1.118/9), which will be included in the extract. The switch element will be provided by the antirepressor, anti-sigma 28 (αS28), which inhibits the transcription of the P28 promoter [89, 225, 217, 219]. A bicistronic construct with αS28 and deGFP downstream of the modified P70 promoter will be constructed. The expression levels of each cistron will be tuned by optimizing the ribosome binding sites and inserting rare codons after the AUG start site.

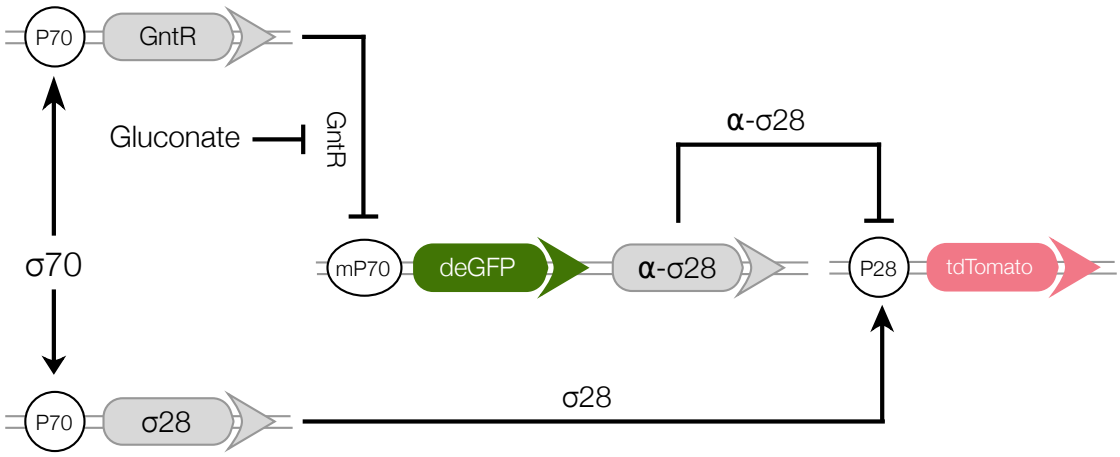


Figure 5.2: Schematic of the glucose sensor for Aim 1. mP70 is the modified P70 promoter with added GntR operator site. αS_{28} is the anti-sigma factor 28 that inhibits the P28 promoter, thereby repressing the expression of tdTomato. In the presence of high levels of gluconate, deGFP protein will be expressed along with αS_{28}. In low gluconate levels, only tdTomato will be expressed.

Insertion of rare codons at this site has been shown to impact translation initiation and expression [186, 206]. Our bicistronic design will follow the strategy used by Hui and coworkers [90]. To confirm the activity of the antirepressor, fluorescence of the two proteins deGFP and tdTomato will be measured; these proteins have non-overlapping spectral profiles and are compatible for simultaneous study [111]. In low glucose levels, the P28 promoter will be on; in high glucose levels, the modified P70 promoter will turn on, expressing αS28 which will repress P28. The selectivity and response time of the switch will also be determined. For all experiments, in order to accelerate the design-build-test cycle, linear gene fragments will be used.

5.1.3 Computational modeling of transcription and translation (TX/TL) with integrated metabolic flux estimation.

The effective biophysical modeling approach built in this body of work, integrated with metabolic flux estimation, will provide additional insights such as energy efficiency of the system and the metabolic consumption of gluconate. Metabolite and amino acid concentrations, quantified by LCMS using the protocol developed in our group [235], can be used to constrain the model. The model results will be validated using mRNA and protein concentration measured using quantitative PCR and fluorescence protein measurements. The integrated model can also provide knowledge of metabolic gluconate consumption pathways. This can, in turn, help develop further engineering strategies such as making appropriate enzyme deletions in the extract preparation process to optimize the gluconate levels for the genetic circuit. The mathematical model can be also used to guide other decisions. For example, the dynamic gene expression profiles estimated by the simulation can aid in experiment design; concentrations of the individual genetic parts can be varied *in silico* to optimize for desired circuit characteristics such as tuning the response time and selectivity of the switch.

BIBLIOGRAPHY

[1] D M F van Aalten, C C DiRusso, J Knudsen, and R K Wierenga. Crystal structure of fadr, a fatty acid-responsive transcription factor with a novel acyl coenzyme a-binding fold. *The EMBO Journal*, 19(19):5167–5177, 2000.

[2] G K Ackers, A D Johnson, and M A Shea. Quantitative model for gene regulation by lambda phage repressor. *Proc Natl Acad Sci U S A*, 79(4):1129–33, Feb 1982.

[3] Katarzyna P Adamala, Daniel A Martin-Alarcon, Katriona R Guthrie-Honea, and Edward S Boyden. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nature chemistry*, 9(5):431, 2017.

[4] Abhinav Adhikari, Michael Vilkhovoy, Sandra Vadhin, Ha Eun Lim, and Jeffrey D. Varner. Effective biophysical modeling of cell free transcription and translation processes. *Frontiers in Bioengineering and Biotechnology*, 8:539081, 2020.

[5] Cem Albayrak, KC Jones, and James R Swartz. Broadening horizons and teaching basic biology through cell-free synthesis of green fluorescent protein in a high school laboratory course. *Journal of Science Education and Technology*, 22(6):963–973, 2013.

[6] Cem Albayrak and James R Swartz. Cell-free co-production of an orthogonal transfer rna activates efficient site-specific non-natural amino acid incorporation. *Nucleic acids research*, 41(11):5949–5963, 2013.

[7] Cem Albayrak and James R Swartz. Direct polymerization of proteins. *ACS synthetic biology*, 3(6):353–362, 2013.

[8] Timothy E Allen and Bernhard Ø Palsson. Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. *J Theor Biol*, 220(1):1–18, Jan 2003.

[9] Jorge Alonso-Gutierrez, Eun-Mi Kim, Tanveer S Batth, Nathan Cho, Qijun Hu, Leanne Jade G Chan, Christopher J Petzold, Nathan J Hillson, Paul D Adams, Jay D Keasling, Hector Garcia Martin, and Taek Soon Lee. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metabolic Engineering*, 28:123–133, 2015.

[10] J C Atlas, E V Nikolaev, S T Browning, and M L Shuler. Incorporating genome-wide dna sequence information into a dynamic whole-cell model of Escherichia coli: application to dna replication. *IET Syst Biol*, 2(5):369–82, Sep 2008.

[11] Shota Atsumi, Taizo Hanai, and James C Liao. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451(7174):86–9, Jan 2008.

[12] JE Bailey. Toward a science of metabolic engineering. *Science*, 252(5013):1668–1675, 1991.

[13] David M Bassen, Michael Vilkhovoy, Mason Minot, Jonathan T Butcher, and Jeffrey D Varner. Jupoets: a constrained multiobjective optimization approach to estimate biochemical model ensembles in the julia programming language. *BMC Syst Biol*, 11(1):10, 01 2017.

[14] Matthias Bechtold, Elisabetta Brenna, Christian Femmer, Francesco G Gatti, Sven Panke, Fabio Parmeggiani, and Alessandro Sacchetti. Biotechnological development of a practical synthesis of ethyl (s)-2-ethoxy-3-(p-methoxyphenyl) propanoate (eehp): over 100-fold productivity increase from yeast whole cells to recombinant isolated enzymes. *Organic Process Research & Development*, 16(2):269–276, 2011.

[15] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.

[16] Igor W Bogorad, Tzu-Shyang Lin, and James C Liao. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature*, 502(7473):693, 2013.

[17] Richard Bonneau, David J Reiss, Paul Shannon, Marc Facciotti, Leroy Hood, Nitin S Baliga, and Vesteinn Thorsson. The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol*, 7(5):R36, 2006.

[18] Henry Borsook. Protein turnover and incorporation of labeled amino acids into tissue proteins in vivo and in vitro. *Physiological Reviews*, 30(2):206–219, 1950.

[19] Marcus E Boyer, James A Stapleton, Jon M Kuchenreuther, Chia-wei Wang, and James R Swartz. Cell-free synthesis and maturation of [fefe] hydrogenases. *Biotechnol Bioeng*, 99(1):59–67, 2008.

[20] Sabine Brantl and E. Gerhart H. Wagner. Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Molecular Microbiology*, 35(6):1469–1482, 2000.

[21] Robert C Brewster, Daniel L Jones, and Rob Phillips. Tuning promoter strength through rna polymerase binding site design in escherichia coli. *PLoS Comput Biol*, 8(12):e1002811, 2012.

[22] Kara A. Calhoun and James R. Swartz. An Economical Method for Cell-Free Protein Synthesis using Glucose and Nucleoside Monophosphates. *Biotechnol Prog*, 21(4):1146–53, 2005.

[23] Kara A Calhoun and James R Swartz. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol Bioeng*, 90(5):606–13, Jun 2005.

[24] Kara A Calhoun and James R Swartz. Total amino acid stabilization during cell-free protein synthesis reactions. *Journal of biotechnology*, 123(2):193–203, 2006.

[25] Erik D Carlson, Rui Gan, C Eric Hodgman, and Michael C Jewett. Cell-free protein synthesis: applications come of age. *Biotechnol Adv*, 30(5):1185–94, 2012.

[26] M Castellanos, D B Wilson, and M L Shuler. A modular minimal cell model: purine and pyrimidine transport and metabolism. *Proc Natl Acad Sci*, 101(17):6681–6, Apr 2004.

[27] Sriram Chandrasekaran and Nathan D Price. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in escherichia coli and mycobacterium tuberculosis. *Proc Natl Acad Sci U S A*, 107(41):17845–50, Oct 2010.

[28] Roger L. Chang, Kathleen Andrews, Donghyuk Kim, Zhanwen Li, Adam Godzik, and Bernhard O. Palsson. Structural systems biology evaluation of metabolic thermotolerance in Escherichia coli. *Science*, 340(6137):1220–1223, 2013.

[29] James Chappell, Melissa K. Takahashi, and Julius B. Lucks. Creating small transcription activating RNAs. *Nature Chemical Biology*, 11(3):214–220, March 2015.

[30] Xixian Chen, Congqiang Zhang, Ruiyang Zou, Gregory Stephanopoulos,

and Heng-Phon Too. In Vitro Metabolic Engineering of Amorpha-4,11-diene Biosynthesis at Enhanced Rate and Specific Yield of Production. *ACS synthetic biology*, 6(9):1691–1700, 2017.

[31] Xixian Chen, Congqiang Zhang, Ruiyang Zou, Kang Zhou, Gregory Stephanopoulos, and Heng Phon Too. Statistical experimental design guided optimization of a one-pot biphasic multienzyme total synthesis of amorpha-4,11-diene. *PloS one*, 8(11):e79650, 2013.

[32] Leonid T Cherney, Maia M Cherney, Craig R Garen, George J Lu, and Michael N G James. Crystal structure of the arginine repressor protein in complex with the dna operator from mycobacterium tuberculosis. *Journal of molecular biology*, 384(5):1330–40, 2008.

[33] Daphne Collias, Ryan Marshall, Scott P Collins, Chase L Beisel, and Vincent Noireaux. An educational module to explore crispr technologies with a cell-free transcription-translation system. *Synthetic Biology*, 4(1):ysz005, 2019.

[34] M W Covert, C H Schilling, and B Ø Palsson. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 213(1):73–88, Nov 2001.

[35] Markus W Covert, Eric M Knight, Jennifer L Reed, Markus J Herrgard, and Bernhard O Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–6, May 2004.

[36] Markus W Covert and Bernhard Ø Palsson. Transcriptional regulation in constraints-based metabolic models of escherichia coli. *J Biol Chem*, 277(31):28058–64, Aug 2002.

[37] Leslie Cuthbertson and Justin R Nodwell. The tetr family of regulators. *Microbiology and molecular biology reviews: MMBR*, 77(3):440–75, 2013.

[38] A. Daddaoua, A. Corral-Lugo, J.-L. Ramos, and Tino Krell. Identification of gntr as regulator of the glucose metabolism in pseudomonas aeruginosa. *Environmental Microbiology*, 19(9):3721–3733, 2017.

[39] D Dai, N Horvath, and J. Varner. Dynamic Sequence Specific Constraint-Based Modeling of Cell-Free Protein Synthesis. *Processes*, 6:132, 2018.

[40] Samar Damiati, Rami Mhanna, Rimantas Kodzius, and Eva-Kathrin

Ehmoser. Cell-free approaches in synthetic biology utilizing microfluidics. *Genes*, 9(3):144, 2018.

[41] Katja Dettmer, Pavel A. Aronov, and Bruce D. Hammock. Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews*, 26(1):51–78, 2007.

[42] Andriy Didovyk, Taishi Tonooka, Lev Tsimring, and Jeff Hasty. Rapid and scalable preparation of bacterial lysates for cell-free gene expression. *ACS synthetic biology*, 6(12):2198–2208, 2017.

[43] Rey P Dimas, Benjamin R Jordan, Xian-Li Jiang, Catherine Martini, Joseph S Glavy, Dustin P Patterson, Faruck Morcos, and Clement T Y Chan. Engineering dna recognition and allosteric response properties of tetr family proteins by using a module-swapping strategy. *Nucleic Acids Research*, 47(16):8913–8925, 2019.

[44] Anne Doerr, Elise de Reus, Pauline van Nies, Mischa van der Haar, Katy Wei, Johannes Kattan, Aljoscha Wahl, and Christophe Danelon. Modelling cell-free RNA and protein synthesis with minimal systems. *Physical Biology*, 16(2):025001, 2019.

[45] M M Domach, S K Leung, R E Cahn, G G Cocks, and M L Shuler. Computer model for glucose-limited growth of a single cell of Escherichia coli b/r-a. *Biotechnol Bioeng*, 26(3):203–16, Mar 1984.

[46] Dijun Du, Hendrik W van Veen, and Ben F Luisi. Assembly and operation of bacterial tripartite multidrug efflux pumps. *Trends in Microbiology*, 23(5):311–319, 2015.

[47] Quentin M Dudley, Kim C Anderson, and Michael C Jewett. Cell-free mixing of escherichia coli crude extracts to prototype and rationally engineer high-titer mevalonate synthesis. *ACS synthetic biology*, 5(12):1578–1588, 2016.

[48] Quentin M Dudley, Ashty S Karim, and Michael C Jewett. Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnology journal*, 10(1):69–82, 2015.

[49] Quentin M Dudley, Connor J Nash, and Michael C Jewett. Cell-free biosynthesis of limonene using enzyme-enriched Escherichia coli lysates. *Synthetic Biology*, 4(1):ysz003, 2019.

[50] Warwick B. Dunn, Alexander Erban, Ralf J M Weber, Darren J. Creek, Marie Brown, Rainer Breitling, Thomas Hankemeier, Royston Goodacre, Steffen Neumann, Joachim Kopka, and Mark R. Viant. Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics : Official journal of the Metabolomic Society*, 9(1):44–66, 2013.

[51] Jeremy S. Edwards and Bernhard O. Palsson. Metabolic flux balance analysis and the in silico analysis of Escherichia coli k-12 gene deletions. *BMC Bioinformatics*, 1(1):1, 2000.

[52] Julia M Flynn, Saskia B Neher, Yong-In Kim, Robert T Sauer, and Tania A Baker. Proteomic Discovery of Cellular Substrates of the ClpXP Protease Reveals Five Classes of ClpX-Recognition Signals. *Molecular Cell*, 11(3):671–683, 2003.

[53] Elena Fossati, Andrew Ekins, Lauren Narcross, Yun Zhu, Jean-Pierre Falgueyret, Guillaume A. W. Beaudoin, Peter J Facchini, and Vincent J. J. Martin. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in saccharomyces cerevisiae. *Nat Commun*, 5, 02 2014.

[54] A G Fredrickson. Formulation of structured growth models. *Biotechnol Bioeng*, 18(10):1481–6, Oct 1976.

[55] Ulysses Amancio de Frias, Greicy Kelly Bonifacio Pereira, María-Eugenia Guazzaroni, and Rafael Silva-Rocha. Boosting secondary metabolite production and discovery through the engineering of novel microbial biosensors. *BioMed Research International*, 2018:1–11, 2018.

[56] Brian R Fritz and Michael C Jewett. The impact of transcriptional tuning on in vitro integrated rrna transcription and ribosome construction. *Nucleic acids research*, 42(10):6774–6785, 2014.

[57] Y Fujita and Y Miwa. Identification of an operator sequence for the Bacillus subtilis gnt operon. *The Journal of biological chemistry*, 264(7):4201–6, 1989.

[58] Ernest F Gale and Joan P Folkes. Effect of nucleic acids on protein synthesis and amino-acid incorporation in disrupted staphylococcal cells. *Nature*, 173:1223–7, 1954.

[59] Jonathan Garamella, Ryan Marshall, Mark Rustad, and Vincent Noireaux. The all e. coli tx-tl toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synth Biol*, 5(4):344–55, Apr 2016.

[60] Jonathan Garamella, Ryan Marshall, Mark Rustad, and Vincent Noireaux. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synthetic Biology*, 5(4):344–355, 2016.

[61] David Garenne, Chase L Beisel, and Vincent Noireaux. Characterization of the all-e. coli transcription-translation system mytxtl by mass spectrometry. *Rapid Commun Mass Spectrom*, 33(11):1036–1048, May 2019.

[62] David Garenne and Vincent Noireaux. Cell-free transcription–translation: engineering biology from the nanometer to the millimeter scale. *Current Opinion in Biotechnology*, 58:19–27, 2019.

[63] V. Georgi, L. Georgi, M. Blechert, M. Bergmeister, M. Zwanzig, D. A. Wüstenhagen, F. F. Bier, E. Jung, and S. Kubick. On-chip automation of cell-free protein synthesis: new opportunities due to a novel reaction mode. *Lab on a Chip*, 16(2):269–281, January 2016.

[64] Belay Gessesse, Takashi Nagaike, Koji Nagata, Yoshihiro Shimizu, and Takuya Ueda. G-protein coupled receptor protein synthesis on a lipid bilayer using a reconstituted cell-free protein synthesis system. *Life*, 8(4):54, 2018.

[65] Leon Glass and Stuart A Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of theoretical Biology*, 39(1):103–129, 1973.

[66] Aaron R. Goerke and James R. Swartz. Development of cell-free protein synthesis platforms for disulfide bonded proteins. *Biotechnol Bioeng*, 99(2):351–367, 2008.

[67] Russell Gould, David M Bassen, Anirikh Chakrabarti, Jeffrey D Varner, and Jonathan Butcher. Population heterogeneity in the epithelial to mesenchymal transition is controlled by nfat and phosphorylated sp1. *PLoS Comput Biol*, 12(12):e1005251, Dec 2016.

[68] Alexander Gräwe, Anna Dreyer, Tobias Vornholt, Ursela Barteczko, Luzia Buchholz, Gila Drews, Uyen Linh Ho, Marta Eva Jackowski, Melissa

Kracht, Janina Lüders, et al. A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. *PLoS one*, 14(3):e0210940, 2019.

[69] Nicole E Gregorio, Max Z Levine, and Javin P Oza. A user's guide to cell-free protein synthesis. *Methods and Protocols*, 2(1):24, 2019.

[70] Chris Grilly, Jesse Stricker, Wyming Lee Pang, Matthew R Bennett, and Jeff Hasty. A synthetic gene network for tuning protein degradation in saccharomyces cerevisiae. *Molecular systems biology*, 3(1), 2007.

[71] Cassandra Guarino and Matthew P DeLisa. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology*, 22(5):596–601, May 2012.

[72] Weihua Guo, Jiayuan Sheng, and Xueyang Feng. Mini-review: In vitro metabolic engineering for biomanufacturing of high-value products. *Computational and Structural Biotechnology Journal*, 15:161 – 167, 2017.

[73] Jan-Karl Guterl, Daniel Garbe, Jörg Carsten, Fabian Steffler, Bettina Sommer, Steven Reiße, Anja Philipp, Martina Haack, Broder Rühmann, Andre Koltermann, Kettling Ulrich, Thomas Brück, and Volker Sieber. Cell-free metabolic engineering: production of chemicals by minimized reaction cascades. *ChemSusChem*, 5(11):2165–2172, 2012.

[74] A. Gyorgy and R. M. Murray. Quantifying resource competition and its effects in the TX-TL system. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 3363–3368, December 2016.

[75] H Hajjaj, P.J Blanc, G Goma, and J François. Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. *FEMS Microbiology Letters*, 164(1):195–200, 1998.

[76] Shogo Hamada, Kenneth Gene Yancey, Yehudah Pardo, Mingzhe Gan, Max Vanatta, Duo An, Yue Hu, Thomas L. Derrien, Roanna Ruiz, Peifeng Liu, Jenny Sabin, and Dan Luo. Dynamic dna material with emergent locomotion behavior powered by artificial metabolism. *Science Robotics*, 4(29), 2019.

[77] Joshua J Hamilton, Vivek Dwivedi, and Jennifer L Reed. Quantitative Assessment of Thermodynamic Constraints on the Solution Space of Genome-Scale Metabolic Models. *Biophys J*, 105(2):512–522, Jul 2013.

[78] Ryan L. Hartman, Jonathan P. McMullen, and Klavs F. Jensen. Deciding Whether To Go with the Flow: Evaluating the Merits of Flow Reactors for Synthesis. *Angewandte Chemie International Edition*, 50(33):7502–7519, 2011.

[79] D J Haydon and J R Guest. A new family of bacterial regulatory proteins. *FEMS Microbiology Letters*, 79(2–3):291–295, 1991.

[80] Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-Based Metabolic Flux Analysis. *Biophys. J*, 92(5):192–1805, Mar 2006.

[81] Jason R Hillebrecht and Shaorong Chong. A comparative study of protein synthesis in in vitro systems: from the prokaryotic reconstituted to the eukaryotic extract-based. *BMC biotechnology*, 8(1):58, 2008.

[82] M B Hoagland, E B Keller, and P C Zamecnik. Enzymatic carboxyl activation of amino acids. *The Journal of biological chemistry*, 218(1):345–58, 1956.

[83] C Eric Hodgman and Michael C Jewett. Cell-free synthetic biology: thinking outside the cell. *Metab Eng*, 14(3):261–9, May 2012.

[84] N Horvath, M Vilkhovoy, J Wayman, K Calhoun, J Swartz, and J. Varner. Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in Escherichia coli. *Metab Eng Comm*, 10:e00113, 2020.

[85] Nicholas Horvath, Michael Vilkhovoy, Joseph A. Wayman, Kara Calhoun, James Swartz, and Jeffrey D. Varner. Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in Escherichia coli. *Metab. Eng. Comm*, 10:e00113, 2020.

[86] Chelsea Y Hu, Jeffrey D Varner, and Julius B Lucks. Generating effective models and parameters for rna genetic circuits. *ACS Synth Biol*, 4(8):914–26, Aug 2015.

[87] Tianjiao Huang, Michael R. Armbruster, John B. Coulton, and James L. Edwards. Chemical tagging in mass spectrometry for systems biology. *Analytical Chemistry*, 91(1):109–125, Jan 2019.

[88] Tianjiao Huang, Maria Toro, Richard Lee, Dawn S. Hui, and James L. Edwards. Multi-functional derivatization of amine, hydroxyl, and carboxy-

late groups for metabolomic investigations of human tissue by electrospray ionization mass spectrometry. *Analyst*, 143:3408–3414, 2018.

[89] Kelly T Hughes and Kalai Mathee. The Anti-sigma Factors. *Annual Review of Microbiology*, 52(1):231–286, 1998.

[90] Chang-Ye Hui, Yan Guo, Lisa Liu, Hao-Qu Zheng, Chao-Xian Gao, and Wen Zhang. Construction of a rfp-laczα bicistronic reporter system and its application in lead biosensing. *PloS one*, 15(1):e0228456, 2020.

[91] Ryo Iizuka, Mai Yamagishi-Shirasaki, and Takashi Funatsu. Kinetic study of de novo chromophore maturation of fluorescent proteins. *Anal Biochem*, 414(2):173–8, Jul 2011.

[92] Keitaro Ishikawa, Kanetomo Sato, Yasufumi Shima, Itaru Urabe, and Tetsuya Yomo. Expression of a cascading genetic network within liposomes. *FEBS letters*, 576(3):387–390, 2004.

[93] Kirsten Jackson, Shouguang Jin, and Z Hugh Fan. Optimization of a miniaturized fluid array device for cell-free protein synthesis. *Biotechnol Bioeng*, 112(12):2459–2467, 2015.

[94] Amber Jannasch, Miroslav Sedlak, and Jiri Adamec. Quantification of pentose phosphate pathway (ppp) metabolites by liquid chromatography-mass spectrometry (lc-ms). *Methods in molecular biology (Clifton, N.J.)*, 708:159–171, 2011.

[95] Thapakorn Jaroentomeechai, Jessica C Stark, Aravind Natarajan, Cameron J Glasscock, Laura E Yates, Karen J Hsu, Milan Mrksich, Michael C Jewett, and Matthew P DeLisa. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nature communications*, 9(1):2686, 2018.

[96] Michael C Jewett, Kara A Calhoun, Alexei Voloshin, Jessica J Wuu, and James R Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Mol Syst Biol*, 4:220, 2008.

[97] Michael C Jewett, Brian R Fritz, Laura E Timmerman, and George M Church. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Molecular Systems Biology*, 9(1):678, 2013.

[98] Michael C Jewett and James R Swartz. Mimicking the Escherichia coli cy-

toplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1):19–26, 2004.

[99] Michael C Jewett and James R Swartz. Substrate replenishment extends protein synthesis with an in vitro translation system designed to mimic the cytoplasm. *Biotechnology and Bioengineering*, 87(4):465–471, 2004.

[100] David K Karig, Sukanya Iyer, Michael L Simpson, and Mitchel J Doktycz. Expression optimization and synthetic gene networks in cell-free systems. *Nucleic acids research*, 40(8):3763–3774, 2011.

[101] Guy Karlebach and Ron Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, 9(10):770–780, 2008.

[102] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, Jul 2012.

[103] Eyal Karzbrun, Jonghyeon Shin, Roy H. Bar-Ziv, and Vincent Noireaux. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Physical Review Letters*, 106(4):048104, 2010.

[104] Eyal Karzbrun, Jonghyeon Shin, Roy H. Bar-Ziv, and Vincent Noireaux. Coarse-Grained Dynamics of Protein Synthesis in a Cell-Free System. *Physical Review Letters*, 106(4), January 2011.

[105] Eyal Karzbrun, Alexandra M Tayar, Vincent Noireaux, and Roy H Bar-Ziv. Programmable on-chip dna compartments as artificial cells. *Science*, 345(6198):829–832, 2014.

[106] GA Kassavetis and MJ Chamberlin. Pausing and termination of transcription within the early region of bacteriophage t7 dna in vitro. *Journal of Biological Chemistry*, 256(6):2777–2786, 1981.

[107] Benjamin B Kaufmann and Alexander van Oudenaarden. Stochastic gene expression: from single molecules to the proteome. *Curr Opin Genet Dev*, 17(2):107–12, Apr 2007.

[108] Takanori Kigawa, Takashi Yabuki, Natsuko Matsuda, Takayoshi Matsuda, Rie Nakajima, Akiko Tanaka, and Shigeyuki Yokoyama. Preparation of

escherichia coli cell extract for highly productive cell-free protein expression. *Journal of structural and functional genomics*, 5(1-2):63–68, 2004.

[109] Dong-Myung Kim and James R. Swartz. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol Bioeng*, 74(4):309–316, 2001.

[110] Tae-Wan Kim, Jung-Won Keum, In-Seok Oh, Cha-Yong Choi, Chang-Gil Park, and Dong-Myung Kim. Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *Journal of biotechnology*, 126(4):554–561, 2006.

[111] Benjamin Kleeman, Andre Olsson, Tess Newkold, Matt Kofron, Monica DeLay, David Hildeman, and H Leighton Grimes. A guide to choosing fluorescent protein combinations for flow cytometric analysis based on spectral overlap. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 93(5):556–562, 2018.

[112] Mathilde Koch, Amir Pandi, Olivier Borkowski, A C Batista, and Jean-Loup Faulon. Custom-made transcriptional biosensors for metabolic engineering. *Current Opinion in Biotechnology*, 59:78–84, 2019.

[113] Borimas Krutsakorn, Kohsuke Honda, Xiaoting Ye, Takashi Imagawa, Xiaoyu Bei, Kenji Okano, and Hisao Ohtake. In vitro production of n-butanol from glucose. *Metabolic engineering*, 20:84–91, 2013.

[114] Muriel Lederman and Geoffrey Zubay. DNA-directed peptide synthesis I. A comparison of T2 and Escherichia coli DNA-directed peptide synthesis in two cell-free systems. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*, 149(1):253–258, 1967.

[115] Sun Bok Lee and James E. Bailey. Genetically structured models forlac promoter–operator function in the Escherichia coli chromosome and in multicopy plasmids: Lac operator function. *Biotechnol Bioeng*, 26(11):1372–1382, 1984.

[116] Roberta Lentini, Michele Forlin, Laura Martini, Cristina Del Bianco, Amy C. Spencer, Domenica Torino, and Sheref S. Mansy. Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synthetic Biology*, (9):482–489, 2013.

[117] Nathan E Lewis, Harish Nagarajan, and Bernhard Ø Palsson. Constrain-

ing the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, 10(4):291–305, Apr 2012.

[118] Jun Li, Liangcai Gu, John Aach, and George M Church. Improved cell-free rna and protein synthesis system. *PLoS one*, 9(9):e106232, 2014.

[119] David V Liu, James F Zawada, and James R Swartz. Streamlining escherichia coli s30 extract preparation for economical cell-free protein synthesis. *Biotechnology Progress*, 21(2):460–465, 2005.

[120] Di Liu, Trent Evans, and Fuzhong Zhang. Applications and advances of metabolite biosensors for metabolic engineering. *Metabolic Engineering*, 31:35–43, 2015.

[121] Yang Liu, Ye Liu, and Meng Wang. Design, optimization and application of small molecule biosensor in metabolic engineering. *Frontiers in microbiology*, 8:2012, 2017.

[122] Yuan Lu, John P Welsh, and James R Swartz. Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. *Proc Natl Acad Sci*, 111(1):125–30, Jan 2014.

[123] Julius B. Lucks, Lei Qi, Vivek K. Mutalik, Denise Wang, and Adam P. Arkin. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc Natl Acad Sci*, 108(21):8617–8622, May 2011.

[124] Julius B Lucks, Lei Qi, Vivek K Mutalik, Denise Wang, and Adam P Arkin. Versatile rna-sensing transcriptional regulators for engineering genetic networks. *Proc Natl Acad Sci U S A*, 108(21):8617–22, May 2011.

[125] Duo Ma, Luhui Shen, Kaiyue Wu, Chris W Diehnelt, and Alexander A Green. Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology*, 3(1):ysy018, 2018.

[126] H Maeda, N Fujita, and A Ishihama. Competition among seven escherichia coli sigma subunits: relative binding affinities to the core rna polymerase. *Nucleic Acids Res*, 28(18):3497–503, Sep 2000.

[127] Ryan Marshall and Vincent Noireaux. Synthetic biology with an all e.

coli txtl system: Quantitative characterization of regulatory elements and gene circuits. In *Synthetic Biology*, pages 61–93. Springer, 2018.

[128] Ryan Marshall and Vincent Noireaux. Quantitative modeling of transcription and translation of an all-e. coli cell-free system. *Sci Rep*, 9(1):11980, Aug 2019.

[129] Rey W Martin, Benjamin J Des Soye, Yong-Chan Kwon, Jennifer Kay, Roderick G Davis, Paul M Thomas, Natalia I Majewska, Cindy X Chen, Ryan D Marcum, Mary Grace Weiss, et al. Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nature communications*, 9(1):1203, 2018.

[130] Laura Martini and Sheref S Mansy. Measuring riboswitch activity in vitro and in artificial cells with purified transcription–translation machinery. In *Artificial Riboswitches*, pages 153–164. Springer, 2014.

[131] Tomoaki Matsuura, Naoki Tanimura, Kazufumi Hosoda, Tetsuya Yomo, and Yoshihiro Shimizu. Reaction dynamics analysis of a reconstituted escherichia coli protein translation system by computational modeling. *Proceedings of the National Academy of Sciences*, 114(8):E1336–E1344, 2017.

[132] J H Matthaei and M W Nirenberg. Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. *Proc Natl Acad Sci*, 47:1580–8, Oct 1961.

[133] Doreen Matthies, Stefan Haberstock, Friederike Joos, Volker Dötsch, Janet Vonck, Frank Bernhard, and Thomas Meier. Cell-free expression and assembly of atp synthase. *Journal of Molecular Biology*, 413(3):593–603, 2011.

[134] Fabio Mavelli, Roberto Marangoni, and Pasquale Stano. A simple protein synthesis model for the pure system operation. *Bulletin of mathematical biology*, 77(6):1185–1212, 2015.

[135] H H McAdams and A Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, 94(3):814–9, Feb 1997.

[136] W R McClure. Rate-limiting steps in rna chain initiation. *Proc Natl Acad Sci U S A*, 77(10):5634–8, Oct 1980.

[137] D. Tyler McQuade and Peter H. Seeberger. Applying Flow Chemistry:

Methods, Materials, and Multistep Synthesis. *The Journal of Organic Chemistry*, 78(13):6384–6389, July 2013.

[138] Jerome T Mettetal, Dale Muzzey, Juan M Pedraza, Ertugrul M Ozbudak, and Alexander van Oudenaarden. Predicting stochastic gene expression dynamics in single cells. *Proc Natl Acad Sci U S A*, 103(19):7304–9, May 2006.

[139] Nathalie Michel-Reydellet, Kara Calhoun, and James Swartz. Amino acid stabilization for cell-free protein synthesis by modification of the escherichia coli genome. *Metabolic engineering*, 6(3):197–203, 2004.

[140] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res*, 38(Database issue):D750–3, Jan 2010.

[141] Seung Eui Min, Kyung-Ho Lee, Seong-Wook Park, Tae Hyeon Yoo, Chan Hee Oh, Ji-Ho Park, Sung Yun Yang, Yong-Sung Kim, and Dong-Myung Kim. Cell-free production and streamlined assay of cytosol-penetrating antibodies. *Biotechnology and bioengineering*, 113(10):2107–12, 2016.

[142] Y Miwa and Y Fujita. Purification and characterization of a repressor for the Bacillus subtilis gnt operon. *The Journal of biological chemistry*, 263(26):13252–7, 1988.

[143] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–53, Nov 2012.

[144] Simon J Moore, James T MacDonald, and Paul S Freemont. Cell-free synthetic biology for in vitro prototype engineering. *Biochemical Society Transactions*, 45(3):785–791, 2017.

[145] Simon J Moore, James T MacDonald, Sarah Wienecke, Alka Ishwarbhai, Argyro Tsipa, Rochelle Aw, Nicolas Kylilis, David J Bell, David W McClymont, Kirsten Jensen, et al. Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proceedings of the National Academy of Sciences*, 115(19):E4340–E4349, 2018.

[146] H J Morowitz. The completeness of molecular biology. *Isr J Med Sci*, 20(9):750–3, Sep 1984.

[147] Max D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, pages 161–174, 1991.

[148] Satoshi Murakami, Rena Matsumoto, and Takashi Kanamori. Constructive approach for synthesis of a functional IgG using a reconstituted cell-free protein synthesis system. *Scientific Reports*, 9(1):671, 2019.

[149] Travis W Murphy, Jiayuan Sheng, Lynette B Naler, Xueyang Feng, and Chang Lu. On-chip manufacturing of synthetic proteins for point-of-care therapeutics. *Microsystems & nanoengineering*, 5(1):1–12, 2019.

[150] Charles E Nakamura and Gregory M Whited. Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology*, 14(5):454 – 459, 2003.

[151] Patrick P Ng, Ming Jia, Kedar G Patel, Joshua D Brody, James R Swartz, Shoshana Levy, and Ronald Levy. A vaccine directed to b cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. *Proc Natl Acad Sci*, 109(36):14526–14531, 2012.

[152] Pamela N. Nge, Chad I. Rogers, and Adam T. Woolley. Advances in Microfluidic Materials, Functions, Integration, and Applications. *Chemical Reviews*, 113(4):2550–2583, April 2013.

[153] Henrike Niederholtmeyer, Zachary Z Sun, Yutaka Hori, Enoch Yeung, Amanda Verpoorte, Richard M Murray, and Sebastian J Maerkl. Rapid cell-free forward engineering of novel genetic ring oscillators. *Elife*, 4:e09771, 2015.

[154] Henrike Niederholtmeyer, Ling Xu, and Sebastian J Maerkl. Real-time mrna measurement during an in vitro transcription and translation reaction using binary probes. *ACS synthetic biology*, 2(8):411–417, 2012.

[155] Alexander Nieß, Jurek Failmezger, Maike Kuschel, Martin Siemann-Herzberg, and Ralf Takors. Experimentally Validated Model Enables Debottlenecking of in Vitro Protein Synthesis and Identifies a Control Shift under in Vivo Conditions. *ACS Synthetic Biology*, 6(10):1913–1921, October 2017.

[156] B Nilsson. Probable in vivo induction of differentiation by retinoic acid of promyelocytes in acute promyelocytic leukaemia. *Br J Haematol*, 57(3):365–71, Jul 1984.

[157] M W Nirenberg and J H Matthaei. The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. *Proc Natl Acad Sci*, 47:1588–602, Oct 1961.

[158] Tatsuya Niwa, Yoshihiro Sasaki, Eri Uemura, Shugo Nakamura, Minato Akiyama, Mitsuru Ando, Shinichi Sawada, Sada-atu Mukai, Takuya Ueda, Hideki Taguchi, et al. Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Scientific reports*, 5:18025, 2015.

[159] Vincent Noireaux and Albert Libchaber. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc Natl Acad Sci*, 101(51):17669–17674, 2004.

[160] Edward J O'Brien, Joshua A Lerman, Roger L Chang, Daniel R Hyduke, and Bernhard Ø Palsson. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Sys. Biol.*, 9(1):693, 2013.

[161] Olotu W Ogonah, Karen M Polizzi, and Daniel G Bracewell. Cell free protein synthesis: a viable option for stratified medicines manufacturing? *Current Opinion in Chemical Engineering*, 18:77–83, 2017.

[162] Paul H Opgenorth, Tyler P Korman, and James U Bowie. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature communications*, 5:4113, 2014.

[163] Paul H Opgenorth, Tyler P Korman, and James U Bowie. A synthetic biochemistry module for production of bio-based chemicals from glucose. *Nature Chemical Biology*, 12(6):393–395, 2016.

[164] Jeffrey D Orth, R M T Fleming, and Bernhard Ø Palsson. Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. *EcoSal Plus*, 4(1), Sep 2010.

[165] Keith Pardee, Alexander A Green, Tom Ferrante, D Ewen Cameron, Ajay DaleyKeyser, Peng Yin, and James J Collins. Paper-based synthetic gene networks. *Cell*, 159(4):940–954, 2014.

[166] Keith Pardee, Alexander A Green, Melissa K Takahashi, Dana Braff, Guillaume Lambert, Jeong Wook Lee, Tom Ferrante, Duo Ma, Nina Donghia, Melina Fan, et al. Rapid, low-cost detection of zika virus using programmable biomolecular components. *Cell*, 165(5):1255–1266, 2016.

[167] Keith Pardee, Shimyn Slomovic, Peter Q Nguyen, Jeong Wook Lee, Nina Donghia, Devin Burrill, Tom Ferrante, Fern R McSorley, Yoshikazu Furuta, Andyna Vernet, Michael Lewandowski, Christopher N Boddy, Neel S Joshi, and James J Collins. Portable, on-demand biomolecular manufacturing. *Cell*, 167(1):248–59.e12, Sep 2016.

[168] Junyoung O Park, Sara A Rubin, Yi-Fan Xu, Daniel Amador-Noguez, Jing Fan, Tomer Shlomi, and Joshua D Rabinowitz. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat Chem Biol*, 12(7):482–9, 07 2016.

[169] Norbert Peekhaus and T Conway. Positive and Negative Transcriptional Regulation of the Escherichia coli Gluconate Regulon Gene gntTby GntR and the Cyclic AMP (cAMP)-cAMP Receptor Protein Complex. *Journal of Bacteriology*, 180(7):1777–1785, 1998.

[170] Tereza Pereira de Souza, Pasquale Stano, and Pier Luigi Luisi. The minimal size of liposome-based model cells brings about a remarkably enhanced entrapment and protein synthesis. *ChemBioChem*, 10(6):1056–1063, 2009.

[171] Jessica G Perez, Jessica C Stark, and Michael C Jewett. Cell-free synthetic biology: engineering beyond the cell. *Cold Spring Harbor perspectives in biology*, 8(12):a023853, 2016.

[172] Tim Prangemeier, François-Xavier Lehr, Rogier M Schoeman, and Heinz Koeppl. Microfluidic platforms for the dynamic characterisation of synthetic circuitry. *Current Opinion in Biotechnology*, 63:167–176, 2020.

[173] Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and T M Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods*, 17(2):147–154, Feb 2020.

[174] JM Pratt. Coupled transcription-translation in prokaryotic cell-free systems. *Transcription and translation: a practical approach.*, pages 179–209, 1984.

[175] Ji Qiu, James R Swartz, and George Georgiou. Expression of Active Human Tissue-Type Plasminogen Activator in Escherichia coli. *Applied and Environmental Microbiology*, 64(12):4891–4896, 1998.

[176] C. Rackauckas and Q. Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *J Open Res Soft*, 5(1):15, 2017.

[177] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–26, Oct 2008.

[178] A Reizer, J Deutscher, M H Saier, and J Reizer. Analysis of the gluconate (gnt) operon of Bacillus subtilis. *Molecular Microbiology*, 5(5):1081–1089, 1991.

[179] Sébastien Rigali, Adeline Derouaux, Fabrizio Giannotta, and Jean Dusart. Subdivision of the helix-turn-helix gntr family of bacterial regulators in fadr, hutc, mocr, and ytra sub-families. *Journal of Biological Chemistry*, 277(15):12507–12515, 2001.

[180] Dae-Kyun Ro, Eric M Paradise, Mario Ouellet, Karl J Fisher, Karyn L Newman, John M Ndungu, Kimberly A Ho, Rachel A Eachus, Timothy S Ham, James Kirby, Michelle C Y Chang, Sydnor T Withers, Yoichiro Shiba, Richmond Sarpong, and Jay D Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–3, Apr 2006.

[181] Gabriel Rosenblum and Barry S. Cooperman. Engine out of the chassis: Cell-free protein synthesis and its uses. *FEBS Letters*, 588(2):261 – 268, 2014. Protein Engineering.

[182] G.J.G. Ruijter and J. Visser. Determination of intermediary metabolites in aspergillus niger. *Journal of Microbiological Methods*, 25(3):295 – 302, 1996.

[183] Ryota Sakamoto, Vincent Noireaux, and Yusuke T Maeda. Anomalous scaling of gene expression in confined cell-free reactions. *Scientific reports*, 8(1):7364, 2018.

[184] Amin S M Salehi, Mark Thomas Smith, Anthony M Bennett, Jacob B Williams, William G Pitt, and Bradley C Bundy. Cell-free protein synthesis of a cytotoxic cancer therapeutic: Onconase production and a just-add-water cell-free system. *Biotechnology Journal*, 11(2):274–281, 2015.

[185] Claudia Sánchez, Juan Carlos Quintero, and Silvia Ochoa. Flux balance analysis in the production of clavulanic acid by Streptomyces clavuligerus. *Biotechnol Prog*, 31(5):1226–1236, 2015.

[186] T Sato, M Terabe, H Watanabe, T Gojobori, C Hori-Takemoto, and K i Miura. Codon and base biases after the initiation codon of the open reading frames in the escherichia coli genome and their influence on the translation efficiency. *Journal of Biochemistry*, 129(6):851–860, 2001.

[187] C H Schilling, D Letscher, and B O Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3):229–48, Apr 2000.

[188] S Schuster, D A Fell, and T Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–32, Mar 2000.

[189] Yoshihiro Shimizu, Akio Inoue, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. Cell-free translation reconstituted with purified components. *Nature biotechnology*, 19(8):751, 2001.

[190] Jonghyeon Shin and Vincent Noireaux. Efficient cell-free expression with the endogenous e. coli rna polymerase and sigma factor 70. *J Biol Eng*, 4:8, Jun 2010.

[191] Jonghyeon Shin and Vincent Noireaux. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of Biological Engineering*, 4(1):8, 2010.

[192] Jonghyeon Shin and Vincent Noireaux. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synthetic Biology*, 1(1):29–41, 2012.

[193] Takehiro Shinoda, Naoko Shinya, Kaori Ito, Yoshiko Ishizuka-Katsura, Noboru Ohsawa, Takaho Terada, Kunio Hirata, Yoshiaki Kawano, Masaki Yamamoto, Taisuke Tomita, et al. Cell-free methods to produce structurally intact mammalian membrane proteins. *Scientific reports*, 6:30442, 2016.

[194] Prashanta Shrestha, Troy Michael Holland, and Bradley Charles Bundy. Streamlined extract preparation for escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. *Biotechniques*, 53(3):163–174, 2012.

[195] Nicolas Sierro, Yuko Makita, Michiel de Hoon, and Kenta Nakai. DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Research*, 36(suppl_1):D93–D96, 2007.

[196] Adam D Silverman, Ashty S Karim, and Michael C Jewett. Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics*, 21(3):151–170, 2019.

[197] Adam D Silverman, Nancy Kelley-Loughnane, Julius B Lucks, and Michael C Jewett. Deconstructing cell-free extract preparation for in vitro activation of transcriptional genetic circuitry. *ACS synthetic biology*, 8(2):403–414, 2018.

[198] Mark Thomas Smith, Anthony M Bennett, Jeremy M Hunt, and Bradley C Bundy. Creating a completely "cell-free" system for protein synthesis. *Biotechnology Progress*, 31(6):1716–1719, 2015.

[199] Jennifer A Sniegowski, Jason W Lappe, Hetal N Patel, Holly A Huffman, and Rebekka M Wachter. Base catalysis of chromophore formation in arg96 and glu222 variants of green fluorescent protein. *J Biol Chem*, 280(28):26248–55, Jul 2005.

[200] Mehran Soltani, Brady R Davis, Hayley Ford, J Andrew D Nelson, and Bradley C Bundy. Reengineering cell-free protein synthesis as a biosensor: Biosensing with transcription, translation, and protein-folding. *Biochemical Engineering Journal*, 138:165–171, 2018.

[201] Mehran Soltani, Brady R. Davis, Hayley Ford, J. Andrew D. Nelson, and Bradley C. Bundy. Reengineering cell-free protein synthesis as a biosensor: Biosensing with transcription, translation, and protein-folding. *Biochemical Engineering Journal*, 138:165–171, 2018.

[202] Hyohak Song and Sang Yup Lee. Production of succinic acid by bacterial fermentation. *Enzyme and Microbial Technology*, 39(3):352 – 361, 2006. The Asia-Pacific Biochemical Engineering Conference (APBioChEC 2005).

[203] A Spirin, V Baranov, L Ryabova, S Ovodov, and Y Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162–1164, 1988.

[204] Jessica C Stark, Thapakorn Jaroentomeechai, Tyler D Moeller, Rachel S Dubner, Karen J Hsu, Taylor C Stevenson, Matthew P DeLisa, and

Michael C Jewett. On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *bioRxiv*, page 681841, 2019.

[205] D.E. Steinmeyer and M.L. Shuler. Structured model for Saccharomyces cerevisiae. *Chem. Eng. Sci.*, 44:2017–2030, 1989.

[206] C Magnus Stenström and Leif A Isaksson. Influences on translation initiation and early elongation by the messenger rna region flanking the initiation codon at the 3 side. *Gene*, 288(1–2):1–8, 2002.

[207] Tobias Stögbauer, Lukas Windhager, Ralf Zimmer, and Joachim O. Rädler. Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integrative Biology*, 4(5):494–501, May 2012.

[208] Tobias Stögbauer, Lukas Windhager, Ralf Zimmer, and Joachim O. Rädler. Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integrative Biology*, 4(5):494–501, 2012.

[209] Inna A Suvorova, Yuri D Korostelev, and Mikhail S Gelfand. Gntr family of bacterial transcription factors and their dna binding motifs: Structure, positioning and co-evolution. *PloS one*, 10(7):e0132618, 2015.

[210] James R. Swartz. Expanding Biological Applications Using Cell-free Metabolic Engineering: An Overview. *Metabolic Engineering*, 50(Berichte der Deutschen Chemischen Gesellschaft 30 1897):156–172, 2018.

[211] James R Swartz. Expanding biological applications using cell-free metabolic engineering: An overview. *Metabolic engineering*, 50:156–172, 2018.

[212] Jim Swartz. Developing cell-free biology for industrial applications. *Journal of Industrial Microbiology & Biotechnology*, 33(7):476–485, 2006.

[213] Kazuyuki Takai, Tatsuya Sawasaki, and Yaeta Endo. Practical cell-free protein synthesis system using purified wheat embryos. *Nature Protocols*, 5, 2001.

[214] R Tapia-Rojo, D Prada-Gracia, J J Mazo, and F Falo. Mesoscopic model for free-energy-landscape analysis of dna sequences. *Phys Rev E Stat Nonlin Soft Matter Phys*, 86(2 Pt 1):021908, Aug 2012.

[215] Rafael Tapia-Rojo, Juan José Mazo, José Ángel Hernández, María Luisa

Peleato, María F Fillat, and Fernando Falo. Mesoscopic model and free energy landscape for protein-dna binding sites: analysis of cyanobacterial promoters. *PLoS Comput Biol*, 10(10):e1003835, Oct 2014.

[216] Ryan Tasseff, Holly A Jensen, Johanna Congleton, David Dai, Katharine V Rogers, Adithya Sagar, Rodica P Bunaciu, Andrew Yen, and Jeffrey D Varner. An effective model of the retinoic acid induced hl-60 differentiation program. *Sci Rep*, 7(1):14327, 10 2017.

[217] Alexandra M. Tayar, Eyal Karzbrun, Vincent Noireaux, and Roy H. Bar-Ziv. Propagating gene expression fronts in a one-dimensional coupled system of artificial cells. *Nature Physics*, 11(12):1037–1041, 2015.

[218] Alexandra M Tayar, Eyal Karzbrun, Vincent Noireaux, and Roy H Bar-Ziv. Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells. *Proc Natl Acad Sci*, 114(44):11609–11614, 2017.

[219] Alexandra M. Tayar, Eyal Karzbrun, Vincent Noireaux, and Roy H. Bar-Ziv. Synchrony and pattern formation of coupled genetic oscillators on a chip of artificial cells. *Proceedings of the National Academy of Sciences*, 114(44):11609–11614, 2017.

[220] Uwe Theobald, Werner Mailinger, Michael Baltes, Manfred Rizzi, and Matthias Reuss. In vivo analysis of metabolic dynamics in saccharomyces cerevisiae : I. experimental observations. *Biotechnol Bioeng*, 55(2):305–316, 1997.

[221] Ines Thiele, Neema Jamshidi, Ronan M. T. Fleming, and Bernhard O. Palsson. Genome-scale reconstruction of Escherichia coli's transcriptional and translational machinery: A knowledge base, its mathematical formulation, and its functional characterization. *PLOS Computational Biology*, 5(3):1–13, 03 2009.

[222] Jean Paul Thiery. Epithelial-mesenchymal transitions in development and pathologies. *Curr Opin Cell Biol*, 15(6):740–6, Dec 2003.

[223] Andrea C. Timm, Peter G. Shankles, Carmen M. Foster, Mitchel J. Doktycz, and Scott T. Retterer. Toward Microfluidic Reactors for Cell-Free Protein Synthesis at the Point-of-Care. *Small*, 12(6):810–817, 2016.

[224] M Tomita, K Hashimoto, K Takahashi, T S Shimizu, Y Matsuzaki, F Miyoshi, K Saito, S Tanida, K Yugi, J C Venter, and C A Hutchison. E-cell:

software environment for whole-cell simulation. *Bioinformatics*, 15(1):72–84, 1999.

[225] Luis Gerardo Treviño-Quintanilla, Julio Augusto Freyre-González, and Irma Martínez-Flores. Anti-sigma factors in e. coli: Common regulatory mechanisms controlling sigma factors availability. *Current genomics*, 14(6):378–87, 2013.

[226] Olga Tsypik, Roman Makitrynskyy, Agnieszka Bera, Lijiang Song, Wolfgang Wohlleben, Victor Fedorenko, and Bohdan Ostash. Role of GntR Family Regulatory Gene SCO1678 in Gluconate Metabolism in Streptomyces coelicolor M145. *BioMed Research International*, 2017:1–9, 2017.

[227] Kelly A. Underwood, James R. Swartz, and Joseph D. Puglisi. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotech. Bioeng.*, 91(4):425–35, 2005.

[228] Meaghan A Valliere, Tyler P Korman, Nicholas B Woodall, Gregory A Khitrov, Robert E Taylor, David Baker, and James U Bowie. A cell-free platform for the prenylation of natural products and application to cannabinoid production. *Nature Communications*, 10(1):565, 2019.

[229] A Varma and B O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl. Environ. Microbiol.*, 60(10):3724–3731, 1994.

[230] Varnerlab. Gene Regulatory Network Generator in Julia (JuGRN), 2.

[231] Varnerlab. Github reposistory for tx/tl model code.

[232] M. Vilkhovoy, M. Minot, and J. D. Varner. Effective dynamic models of metabolic networks. *IEEE Life Sciences Letters*, 2(4):51–54, Dec 2016.

[233] M. Vilkhovoy, M. Minot, and J. D. Varner. Effective dynamic models of metabolic networks. *IEEE Life Sciences Letters*, 2(4):51–54, Dec 2016.

[234] Michael Vilkhovoy, Abhinav Adhikari, Sandra Vadhin, and Jeffrey D Varner. The evolution of cell free biomanufacturing. *Processes*, 8(6):675, 2020.

[235] Michael Vilkhovoy, David Dai, Sandra Vadhin, Abhinav Adhikari, and Jeffrey D Varner. Absolute Quantification of Cell-Free Protein Synthesis

Metabolism by Reversed-Phase Liquid Chromatography-Mass Spectrometry. *Journal of visualized experiments : JoVE*, (152), 2019.

[236] Michael Vilkhovoy, Nicholas Horvath, Che-Hsiao Shih, Joseph A Wayman, Kara Calhoun, James Swartz, and Jeffrey D Varner. Sequence Specific Modeling of E. coli Cell-Free Protein Synthesis. *ACS Synthetic Biology*, 7(8):1844–1857, 2018.

[237] Michael Vilkhovoy, Nicholas Horvath, Che-Hsiao Shih, Joseph A Wayman, Kara Calhoun, James Swartz, and Jeffrey D Varner. Sequence Specific Modeling of *E. coli* Cell-Free Protein Synthesis. *ACS Synth Biol*, 7(8):1844–1857, 08 2018.

[238] Alexei M Voloshin and James R Swartz. Efficient and scalable method for scaling up cell free protein synthesis in batch mode. *Biotechnol Bioeng*, 91(4):516–21, Aug 2005.

[239] Peter L. Voyvodic and Jerome Bonnet. Cell-free biosensors for biomedical applications. *Current Opinion in Biomedical Engineering*, 13:9–15, 2019.

[240] Xiaoqiang Wang, Karolina Corin, Philipp Baaske, Christoph J Wienken, Moran Jerabek-Willemsen, Stefan Duhr, Dieter Braun, and Shuguang Zhang. Peptide surfactants for cell-free production of functional g protein-coupled receptors. *Proc Natl Acad Sci*, 108(22):9049–9054, 2011.

[241] Joseph A. Wayman, Adithya Sagar, and Jeffrey D. Varner. Dynamic modeling of cell-free biochemical networks using effective kinetic models. *Processes*, 3(1):138, 2015.

[242] John P Welsh, Yuan Lu, Xiao-Song He, Harry B Greenberg, and James R Swartz. Cell-free production of trimeric influenza hemagglutinin head domain proteins as vaccine antigens. *Biotechnol Bioeng*, 109(12):2962–2969, 2012.

[243] Wolfgang Wiechert. 13C Metabolic Flux Analysis. *Metabol. Eng.*, 3(3):195 – 206, 2001.

[244] T Winnick. Incorporation of labeled amino acids into the protein of embryonic and tumor tissue homogenates. 9(1):247, 1950.

[245] T Winnick. Studies on the mechanism of protein synthesis in embryonic and tumor tissues. i. evidence relating to the incorporation of la-

beled amino acids into protein structure in homogenates. *Arch Biochem*, 27(1):65–74, Jun 1950.

[246] T Winnick. Studies on the mechanism of protein synthesis in embryonic and tumor tissues. ii. inactivation of fetal rat liver homogenates by dialysis, and reactivation by the adenylic acid system. *Arch Biochem*, 28(3):338–47, Oct 1950.

[247] P Wu, N G Ray, and M L Shuler. A single-cell model for cho cells. *Ann N Y Acad Sci*, 665:152–87, Oct 1992.

[248] Wen-Chu Yang, Miroslav Sedlak, Fred E. Regnier, Nathan Mosier, Nancy Ho, and Jiri Adamec. Simultaneous quantification of metabolites involved in central carbon and energy metabolism using reversed-phase liquid chromatography-mass spectrometry and in vitro 13c labeling. *Analytical Chemistry*, 80(24):9508–9516, Dec 2008.

[249] Maaruthy Yelleswarapu, Ardjan J van der Linden, Bob van Sluijs, Pascal A Pieters, Emilien Dubuc, Tom F A de Greef, and Wilhelm T S Huck. Sigma factor-mediated tuning of bacterial cell-free synthetic genetic oscillators. *ACS Synthetic Biology*, 7(12):2879–2887, 2018.

[250] Tong Yi, Hye Jin Lim, So Jeong Lee, Kyung-Ho Lee, and Dong-Myung Kim. Synthesis of (R,R)-2,3-butanediol from starch in a hybrid cell-free reaction system. *Journal of Industrial and Engineering Chemistry*, 67:231–235, 2018.

[251] Harry Yim, Robert Haselbeck, Wei Niu, Catherine Pujol-Baxley, Anthony Burgard, Jeff Boldt, Julia Khandurina, John D Trawick, Robin E Osterhout, Rosary Stephen, Jazell Estadilla, Sy Teisan, H Brett Schreyer, Stefan Andrae, Tae Hoon Yang, Sang Yup Lee, Mark J Burk, and Stephen Van Dien. Metabolic engineering of escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol*, 7(7):445–452, 07 2011.

[252] Gang Yin, Eudean D Garces, Junhao Yang, Juan Zhang, Cuong Tran, Alexander R Steiner, Christine Roos, Sunil Bajad, Susan Hudak, Kalyani Penta, James Zawada, Sonia Pollitt, and Christopher J Murray. Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *mAbs*, 4(2):217–25, 2012.

[253] Gang Yin and James R Swartz. Enhancing multiple disulfide bonded protein folding in a cell-free system. *Biotechnology and Bioengineering*, 86(2):188–195, 2004.

[254] K Yoshida, H Ohmori, Y Miwa, and Y Fujita. Bacillus subtilis gnt repressor mutants that diminish gluconate-binding ability. *Journal of bacteriology*, 177(16):4813–4816, 1995.

[255] K-I. Yoshida, Y. Miwa, H. Ohmori, and Y. Fujita. Analysis of an insertional operator mutation (gntOi) that affects the expression level of theBacillus subtilis gnt operon, and characterization ofgntOi suppressor mutations. *Molecular and General Genetics MGG*, 248(5):583–591, 1995.

[256] Ken-Ichi Yoshida, Yasutaro Fujita, and Akinori Sarai. Missense mutations in the bacillus subtilis gnt repressor that diminish operator binding ability. *Journal of Molecular Biology*, 231(2):167–174, 1993.

[257] Ke Yue, Yiyong Zhu, and Lei Kai. Cell-free protein synthesis: chassis toward the minimal cell. *Cells*, 8(4):315, 2019.

[258] James F Zawada, Gang Yin, Alexander R Steiner, Junhao Yang, Alpana Naresh, Sushmita M Roy, Daniel S Gold, Henry G Heinsohn, and Christopher J Murray. Microscale to manufacturing scale-up of cell-free cytokine production–a new approach for shortening protein production development timelines. *Biotechnol Bioeng*, 108(7):1570–8, Jul 2011.

[259] Ying Zhang, Ines Thiele, Dana Weekes, Zhanwen Li, Lukasz Jaroszewski, Krzysztof Ginalski, Ashley M. Deacon, John Wooley, Scott A. Lesley, Ian A. Wilson, Bernhard Palsson, Andrei Osterman, and Adam Godzik. Three-dimensional structural view of the central metabolic network of thermotoga maritima. *Science*, 325(5947):1544–1549, 2009.

[260] Xiaoyu Zhou, Han Wu, Miao Cui, Sze Nga Lai, and Bo Zheng. Long-lived protein expression in hydrogel particles: towards artificial cells. *Chemical science*, 9(18):4275–4279, 2018.

[261] Ying Zhou, Haruichi Asahara, Eric A Gaucher, and Shaorong Chong. Reconstitution of translation from thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic acids research*, 40(16):7932–7945, 2012.

[262] Zhiguang Zhu, Tsz Kin Tam, Fangfang Sun, Chun You, and Y-H Percival Zhang. A high-energy-density sugar biobattery based on a synthetic enzymatic pathway. *Nature communications*, 5:3026, 2014.

[263] Geoffrey Zubay. In vitro synthesis of protein in microbial systems. *Annual review of genetics*, 7(1):267–287, 1973.