APPLICATION OF CELL-FREE SYSTEMS FOR THE DESIGN OF A GLUCOSE-RESPONSIVE BIOSENSOR AND FOR STUDYING THE EFFECTS OF KEY SARS-COV-2 PROTEINS

A Dissertation

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

Anirudh Murali Narayanan

December 2021

ABSTRACT

Cell-free platforms have been key in the rapid advancement of synthetic biology. They enable the design of metabolic pathways for the production of any desired proteins. Cell-free systems also present many advantages over traditional in vivo systems such as higher synthesis rates, direct manipulation of the chemical environment, and the ability to even produce proteins toxic to cells. In this study, we made use of the PURE cell-free system to construct a genetic circuit to sense glucose by use of a transcription factor, GntR. Towards this aim, we performed tests of its repression and de-repression characteristics on a reporter with a promoter construct containing the GntR operator site. The Pareto Optimal Ensemble Technique was used to determine the parameters of the model for the simulation of this circuit. This work helps to provide a cell-free alternative for rapid point-of-care detection of glucose for diabetics. Additionally, in light of the recent Covid-19 pandemic, we made use of the myTXTL cell-free system to study some key SARS-CoV-2 proteins. We synthesized the viral host translation inhibitor protein, nsp1 in cell-free and tested its effectiveness to inhibit the translation of a reporter in the same cell-free system. Further, we developed a model to simulate the effect of the RNA-dependent RNA Polymerase (RdRp) on viral protein replication in the myTXTL cell-free system. This work paves a way for simulating the effect of RdRp on viral infection in host cells.

BIOGRAPHICAL SKETCH

Anirudh is from Chennai, India. He graduated with a Bachelor of Technology in Chemical Engineering from Anna University in 2019. In the fall of that year, he joined the Master of Science Program at the Robert Frederick Smith School of Chemical and Biomolecular Engineering at Cornell University and started his research with Dr. Jeffrey D. Varner. It is difficult to find an Anirudh in the wild, but he has occasionally been spotted haunting the halls of Olin or Kimball after midnight.

This work is dedicated to my awesome brother, Vignesh.

ACKNOWLEDGEMENTS

I would like to thank Dr. Jeffrey D. Varner for giving me the opportunity to work in his research group as well as the guidance and support he has given me over the past two years. My gratitude also goes out to Dr. Matthew DeLisa for his insightful comments and his amazing courses which have been instrumental in shaping my interests. I am grateful for the friendship and support I received from all VarnerLab members. In particular, I would like to thank Abhinav Adhikari for mentoring and training me in pretty much everything I can do in the wet lab. Thank you Alagappan, Sachin, Vashist, and Vivek for those weekly game nights that often ran into the mornings. They were the highlights of each week during the Covid-19 pandemic. I would also like to thank the amazing friends I have had in Ithaca, particularly Avinash, Arvind, Abhaiguru, and Harish for being such amazing housemates. Most importantly, I would like to express my gratitude to my parents for encouraging and supporting me all these years. Finally, I would like to thank my brothers, sisters, grandparents, great-grandparents, and every uncle and aunt of my amazing family for always being there for me.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Cell-free protein synthesis

Cell-free systems (CFS) have recently evolved into key platforms for synthetic biology applications. Cell-free systems present many advantages over traditional in vivo systems. Since cell-free systems don't require cellular growth and maintenance, they can allow for higher synthesis rates by allocating more resources towards production of the desired product. The lack of cellular growth also means that they can be used for the production of proteins that are otherwise difficult to produce due to their toxicity to cells. It is also possible to directly observe and manipulate the chemical environment of cell-free systems allowing for rapid tuning of the reaction conditions [46, 92, 179].

This chapter is adapted with permission from the study of my colleagues Vilkhovoy et al. at the VarnerLab [179]. Cell-free systems have historically been used as investigative tools for Molecular Biology research. Some of the first examples of their use were in the 1950s by Borsook [15] and Winnick [190] to study the incorporation of amino acids into proteins. The work of Gale et al. to study the effect of nucleic acids in protein synthesis made use of *Staphylococcus aureus* extract [40]. Even the famous Nirenberg and Matthaei experiments in the early 1960s to discover the genetic code was performed using *E. coli* cell-free extracts [100, 117].

One of the first precursors to modern cell-free transcription and translation (TXTL) platforms was developed by Lederman and Zubay [84] who developed

a coupled transcription-translation bacterial extract that allowed DNA to be used as a template to directly synthesize proteins. In 1988, Spirin et al. developed a continuous cell-free system with a continuous flow of the feeding buffer and a continuous removal of the protein product. They were able to run their system for tens of hours [155]. The Swartz lab improved the energy efficiency of E.coli based CFPS by generating ATP with substrate-level phosphorylation [77]. And this was further improved by oxidative phosphorylation in the Cytomim system that they developed [67–69]. Then the Noireaux lab developed the myTXTL platform that couples ATP regeneration with inorganic phosphate recycling thereby further boosting the protein production [42].

Cell-free platforms can also be used to design synthetic genetic circuits to control gene expression. Bacteriophage-encoded RNA polymerases (RNAP) such as T7, T3, and SP6 RNAP are commonly used in CFPS for their higher transcription efficiency [118, 143]. The use of bacterial regulatory elements based on the sigma factor family has allowed for the implementation of multi-layer genetic cascades in which the protein produced in each stage is the input that is required to promote or inhibit the following stage [2, 42, 118, 143, 144].

1.2 Types of Cell-free systems

Cell-free systems can be broadly classified into two major classes: crude cell lysate-based systems and reconstituted systems. Cell lysate-based systems have been in use for the major part of the history of cell-free systems. These are prepared by the lysis of cells where the cell's transcription and translation machinery is retained while cellular debris and chromosomal DNA are discarded.

This is followed by the addition of necessary components such as amino acids, energy sources, salts, and buffers. They can be prepared from a variety of cells such as cells from bacteriae, protozoa, plants, insects, and mammals. *E. coli* , *S. cerevisiae*, HeLa, rabbit reticulocytes, wheat germ, and High Five insect cells are commonly used cells for the preparation of such cell-free extracts [9,35,42,127,132,149,193].

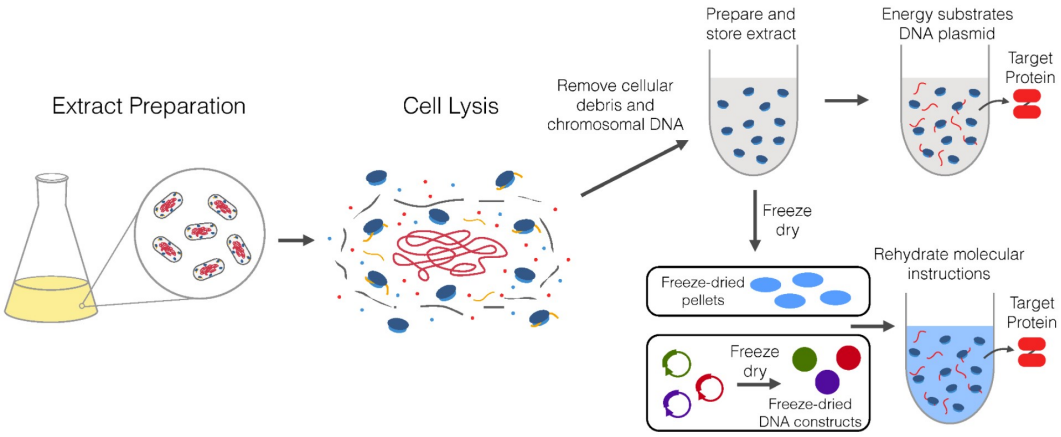


Figure 1.1: Cell-free system based on cell lysate. The Cell-free lysate is prepared by the lysis of cells followed by removal of cellular debris and chromosome DNA. The energy sources, amino acids, nucleotides, and cofactors are added to this to prepare the cell-free extract. The template DNA can be added to it immediately to express the target protein. Alternatively, the cell-free extract can be freeze-dried as pellets for storage or transportation along with lyophilized DNA constructs. These can be rehydrated through the simple addition of water for the synthesis of target proteins. [179]

On the other hand, reconstituted systems are well defined, and are prepared using a "bottom-up" approach where the factors essential for protein synthesis such as the purified enzymes, tRNAs, ribosomes (either prokaryotic or eukaryotic), amino acids, NTPs, and other energy molecules are reconstituted into a system. The PURE (Protein synthesis Using Recombinant Elements) system developed by Shimizu et al. was one of the first reconstituted cell-free systems [142].

Both classes of cell-free systems tend to have their own set of benefits and drawbacks. Crude cell extracts tend to be less expensive and have higher yields than reconstituted systems [46]. They also offer more complex metabolic capabilities that can be exploited for energy regeneration, that allow for larger-scale reactions and extend the duration of protein synthesis [56, 160].

Reconstituted systems also have their own advantages. For one, since the reconstituted systems are prepared from individual components, their exact composition is known and they can be used for studying biological processes including protein expression and folding in the context of a completely defined system. For instance, Li et al. showed that the efficiency of the PURE system could be improved up to 5 fold by adding or adjusting a variety of factors that affect transcription and translation, such as elongation factors, ribosome recycling factors, release factors, chaperones, BSA and tRNAs. showed that the efficiency of protein synthesis was limited by translation elongation capacity, ribosome release, and ribosome recycling [90]. Reconstituted systems such as the PURE system can also be reconstituted in such a way that they do not contain proteases and nucleases, to further improve protein production [142, 160].

1.3 Applications of Cell-free platforms

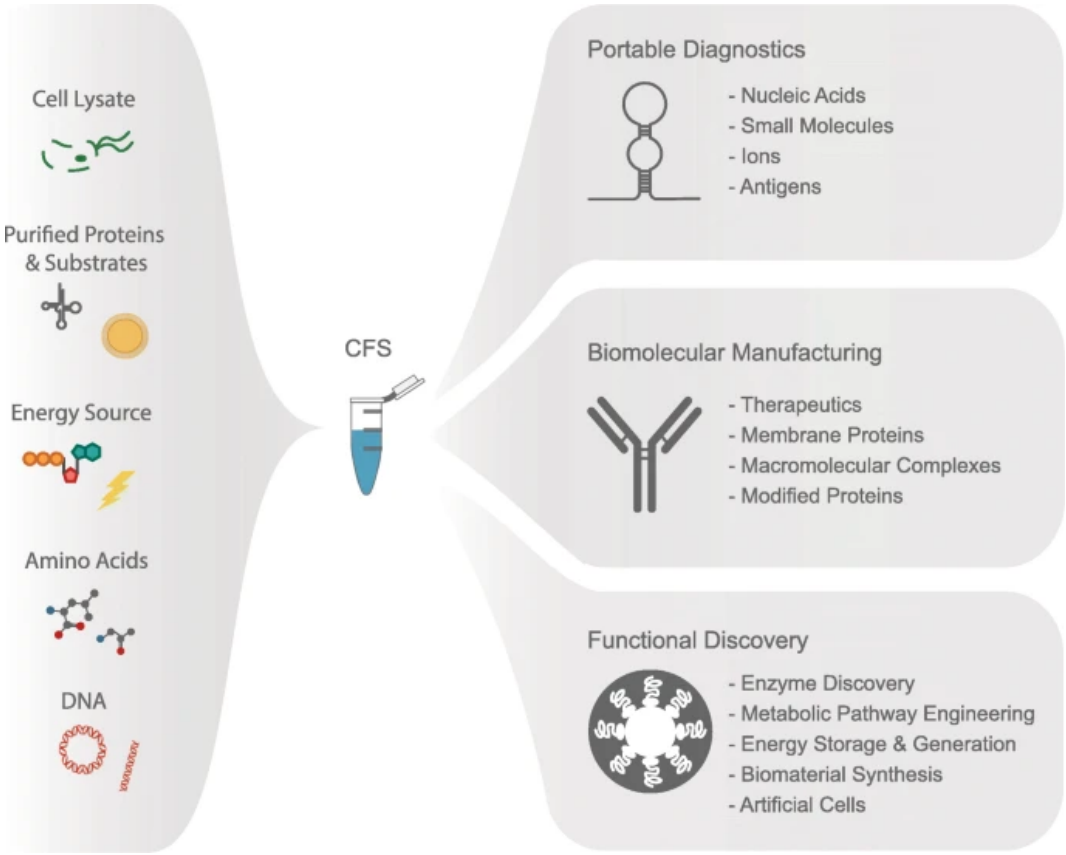


Figure 1.2: Cell-free systems and their applications. The left-hand side of the image indicates the components needed for a cell-free system and the right-hand side indicates the various fields in which cell-free systems can be applied. [170]

Cell-free systems have found many roles in synthetic biology ranging from portable diagnostics to fundamental discovery and prototyping. Cell-free systems have been very useful for the development of biosensors because of their unique advantages over cell-based sensors. They can be used to detect cell wall-impermeable or cytotoxic analytes, and they are immune to issues like mutations and plasmid loss [149]. Cell-free biosensors can be lyophilized (freeze-dried) which allows them to maintain their activity for months for use in

portable diagnostics. They can then be transported to the desired location, rehydrated with water for field use [121,160]. Low-cost cell-free paper-based biosensors have been utilized for the detection of viruses such as norovirus, Ebola virus, and Zika virus [97, 121, 122]. Hamada et al. made use of cell-free platforms for the bottom-up construction of dynamic biomaterials with emergent locomotion behavior powered by artificial metabolism. These could have applications in pathogen detection, and as scaffolds for hybrid nanomaterials [51].

Cell-free systems have a huge scope for both small-scale prototyping of biological processes and larger-scale bioengineering efforts. Because of their ability to allow for direct manipulation of the chemical environment, prototyping is more efficient in cell-free systems. They can also be used to prototype individual genetic parts such as promoters or complex genetic designs in vitro before implementing them in vivo. Moreover, the genetic constructs need not be assembled into plasmids as cell-free systems even allow for the use of linear DNA [149]. Cell-free systems can also be used to analyze how individual genetic parts function together in synthetic genetic circuits by considering each individual part as a logic gate. A number of cell-free genetic circuits have been assembled and prototyped using this strategy including sigma factor-based cascades, RNA transcriptional cascades, RNA single input modules, and feed-forward/feedback loops [2, 42, 118, 162, 163].

Developments in experimental setup and analysis techniques such as the use of real-time fluorescent probes for mRNA measurement, automated liquid-handling robotic systems for rapid screening of cell-free reactions, microfluidic devices, and cell-sized droplet-based expression have all improved the efficiency of prototyping in cell-free systems [25, 106, 107, 115, 133, 149].

1.3.1 Production of Biologics and Specialized Proteins

CFPS can be utilized for the production of biologics and other specialized proteins such as membrane proteins [44,93,113,145,160,185]. The ease of rapid parallel assay and screening with cell-free systems makes them important for the production of libraries for protein engineering and post-genomic research [164]. Recently, cell-free systems have also been used to achieve proteins with complex post-translational modifications such as N-linked glycoproteins [47]. A single-pot glycoprotein synthesis system has also been developed by Jaroentomeechai et al. that allows for on-demand biomanufacturing of glycoproteins [65]. Even proteins incorporating noncanonical amino acids have been synthesized using CFPS systems [6,99].

By constructing cell-free systems in microfluidic devices, we can achieve point-of-care (POC) production of proteins. Continuous flow microfluidic reactors offer more precise control over mixing when compared to batch reactors [53,102]. Georgi et al. have developed an automated, microfluidic CFPS reactor called the TRITT (Transcription - RNA Immobilization and Transfer - Translation) platform that runs both transcription and translation reactions in separate compartments [43]. The quasi-continuous transfer of mRNA to the translation chamber allows for longer CFPS reactions and better protein yield. Microfluidic platforms can also be constructed with integrated purification methods such as dialysis and affinity chromatography [107,114].

1.3.2 Metabolic Engineering

Cell-free systems possess many advantages over in vivo systems for metabolic engineering applications. The cells in vivo systems have their own objectives such as growth or maintenance, which uses up more resources and drives metabolic flux away from the desired pathways making it challenging to achieve high flux in those pathways. On the other hand, in cell-free systems, the allocation of resources to the production of desired products can be maximized [48]. The complexity of living cells also makes computational modeling and optimization of metabolic flux difficult [32]. Cell-free systems, on the other hand, can be accurately modeled and the reaction environment can be directly manipulated and tuned according to the bio-synthetic needs.

Metabolic engineering in cell-free systems has been used to increase flux through enzymatic pathways and improve product yield for the production of high-value small molecule products such as cannabinoids, polyhydroxybutyrate bioplastic, isobutanol, ethanol, limonene, etc. [8,33,49,82,120,174]. It has also been used to handle bottlenecks in CFPS such as the need for energy and cofactor regeneration in the system [14,18,68,77].

1.4 Mathematical Modeling of Cell-free systems

There have been several mathematical models of CFPS. Most of these models focus on the TXTL processes and are mostly systems of ordinary differential equations (ODEs) based on saturation or Hill-like equations.

Karzbrun et al. derived a coarse-grained model comprising the entire en-

dogenous transcription, translation, as well as mRNA and protein degradation machinery for an *E.coli* cell-free extract [74]. To simplify calculations, this model was based on four enzymes and ten relevant rate constants. Transcription and translation processes were assumed to follow Michaelis–Menten kinetics. They assumed that the protein expression was governed by zeroth-order degradation and noted that the expression followed a sharp transition from undetectable levels to constant-rate accumulation, without reaching steady state. However, their study only focused on the first hour of the reaction as the protein synthesis rate of their system began to exponentially decay after that. This sudden decay was assumed to be due to resource depletion and waste accumulation.

Stögbauer et al. developed a model that accounted for resource consumption and degradation [157]. The model predicts that the protein yield depends on the experiment timing in addition to template DNA concentration. They attempted to use Hill functions to better predict the observed saturation effects of mRNA and protein, but the optimized Hill coefficients were close to one, effectively reducing the Hill equation to Michaelis–Menten equation. They identified ribosomal degradation as the cause for the cessation of cell-free protein synthesis in the specific reconstituted system that they used.

More recently, Neiß et al. developed a comprehensive experimentally validated model that was able to identify the major limiting factors of cell-free protein synthesis - the supply of ternary complexes consisting of EFTu and tRNA [116]. This model used an unusual hybrid black-box approach: Although the translation processes were well detailed, the transcription processes were simplified. The entire model was a large system of differential algebraic equations - a system of eight algebraic equations and over 400 ODEs. Gyorgy and

Murry's model explicitly included competition for shared resources in genetic networks [50]. This model also predicted possible product concentrations in multiple-protein expression systems. The authors illustrated that resource competition was a key consideration in genetic circuit design.

Genetic circuits based on regulation by RNA have also been explored and mathematical models for such systems have been developed for cell-free platforms. Transcriptional regulatory RNAs are of interest because they bypass the need for regulatory proteins and they have been used to create various logic gates and cascades [17, 21, 60, 94]. Hu et al. developed an experimentally validated model of a synthetic RNA circuit that contained 8 ODEs and 13 previously unknown parameters which were estimated from sensitivity analysis [60]. Although these models of transcription, translation, resource competition, and gene regulatory circuits have provided useful information for optimizing cell-free biomanufacturing, more sophisticated models integrating metabolic pathways are necessary.

1.4.1 Metabolic Modeling

Traditional approaches to metabolic modeling, which were first developed to describe living cells, could also be applied to cell-free systems. Mechanistic in vivo metabolic models arose from the desire to predict microbial phenotypes resulting from changes in intracellular or extracellular states [38]. The Shuler lab constructed large-scale, dynamic metabolic models for single-cells that incorporated multiple regulated catabolic and anabolic pathways constrained by experimentally determined kinetic parameters [31, 156, 192], minimal cell archi-

tectures [19], and DNA sequence-based whole-cell models of *E. coli* [10]. Next-generation metabolic models described cellular processes such as RNA synthesis, chromosome synthesis, and regulated catabolic and macromolecular synthesis pathways in detail using ODEs [171]. Karr et al. even developed a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions inside the cell.

However, traditional metabolic modeling approaches are often complex and nonlinear and require the estimation of a large number of unknown parameters. This is a difficult process because of the inherent noisiness of biological data and the computational complexity involved in the repeated solving of the model equations. To overcome such obstacles, constraint-based methods were developed to describe metabolic networks with only a limited need for kinetic parameters [175]. Constraint-based approaches such as Flux Balance Analysis (FBA) have been used for the stoichiometric reconstructions of metabolism [88]. FBA, Metabolic Flux Analysis (MFA), and convex network decomposition approaches such as elementary modes and extreme pathways have been developed to model intracellular metabolism using the biochemical stoichiometry and other constraints such as thermodynamic feasibility under pseudo-steady-state conditions [52, 55, 137, 139, 189]. Such constraint-based approaches use linear programming to predict productivity, mutant behavior, and growth phenotypes for biochemical networks of even genome-scale networks [34, 119, 135, 175]. The objective function used in such constraint-based approaches in a cell-free system is the maximization of the desired protein yield.

These stoichiometric reconstructions have been expanded to even include

the metabolic demands for protein synthesis. One of the earliest such models were the sequence-specific constraint-based models developed by Allen and Palsson, which depended on the DNA and protein sequences of interest and the TXTL processes were integrated with metabolism [7]. Since then, sequence-specific constraint-based models have been expanded to the genome-scale with detailed descriptions of gene expression (ME-model) and protein structures (GEM-PRO) [7, 20, 88, 119, 201]. These developments have greatly expanded the scope of stoichiometric reconstructions. Constraint-based methods can be potentially used to predict non-intuitive strategies to optimize the interaction between metabolism and gene expression in cell-free platforms. Thus, the use of integrated constraint-based models for cell-free optimization studies is a promising future research direction.

1.4.2 Integrated cell-free models

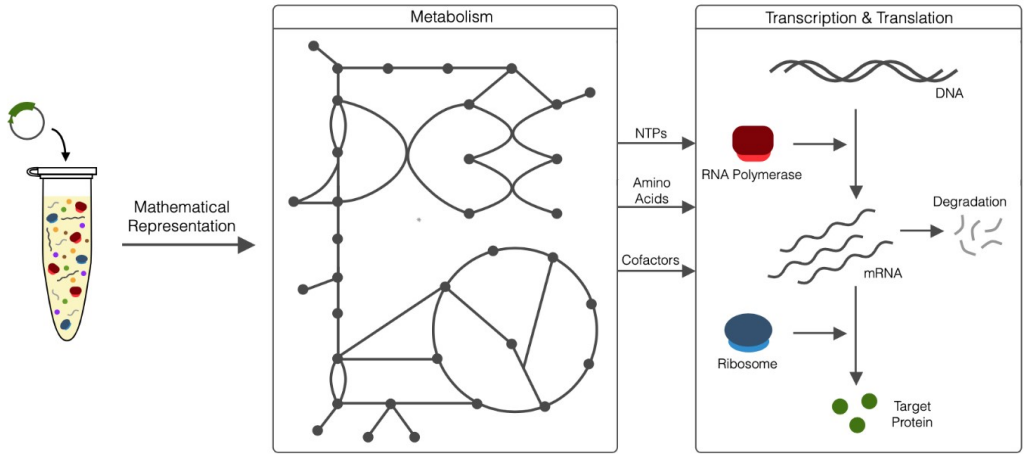


Figure 1.3: The integration of metabolism with transcription and translation (TXTL) processes: The TXTL processes utilize the macromolecular precursors from metabolism (such as NTPs, amino acids, and cofactors) for gene expression. The integrated framework can be represented as a stoichiometric matrix of metabolites participating in certain reactions, along with a description of the metabolic demands for protein expression. By applying a pseudo-steady-state assumption, making use of various constraints and an objective function, the metabolic fluxes in the system can be estimated. In the case of cell-free systems, the objective function is the maximization of target protein yield. [177]

Although CFPS also depends on central carbon metabolism and other metabolic pathways and doesn't just rely on TXTL processes, very few mathematical models have managed to integrate cell-free TXTL processes with metabolic pathways [24, 59, 181]. Wayman et al. developed a hybrid cell-free modeling approach of Wayman that integrated kinetic modeling with a rule-based description of allosteric control [186]. Horvath et al. developed an ensemble of dynamic *E. coli* CFPS models on top of this, using parameters estimated from measurements of metabolite, amino acid, and protein concentrations from CFPS reactions conducted using the PANOx-SP cell-free system [59]. By simulating reaction group knockouts, they suggested that cell-free metabolism and protein

synthesis were strongly coupled with oxidative phosphorylation and glycolytic flux. Vilkhovoy et al. developed an experimentally validated constraint-based model of CFPS [181] to circumvent computationally expensive parameter estimation. This model integrated the expression of a model protein product with the supply of metabolic precursors and energy (Figure 1.3). This model coupled TXTL processes with the 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 was also able to describe protein expression in the myTXTL system using only a limited number of experimentally derived parameters. Taken together, the incorporation of complex metabolism with genetic regulatory networks using constraint-based modeling is a promising approach to simulate cell-free systems.

CHAPTER 2

DEVELOPMENT OF A GLUCONATE-RESPONSIVE TRANSCRIPTION FACTOR-BASED GENETIC CIRCUIT IN A RECONSTITUTED CELL-FREE SYSTEM

2.1 Introduction

Diabetes mellitus is a group of metabolic diseases that can be characterized by chronic hyperglycemia occurring due to imperfect insulin synthesis, insulin action, or both [76]. Diabetes is increasing at an alarming rate globally. According to International Diabetes Federation, about 463 million adults were living with diabetes in 2019 and this number is expected to grow to 700 million by 2045 [1]. There are two main classes of the disease: (i) Type 1 diabetes occurs due to the destruction of β cells of the pancreas causing the body to produce little to no insulin [26, 30]. (ii) Type 2 diabetes is the more prevalent one where the body doesn't utilize the insulin well and may also not produce insulin [76].

Insulin is a hormone necessary to prevent blood sugar levels from rising too high (hyperglycemia) by allowing adipose tissues to take up glucose. Glucagon is a counter-regulatory hormone for insulin that stops blood sugar levels from dropping too low (hypoglycemia) by breaking down glycogen to produce glucose [70]. Cell-free systems can be engineered to create an input-responsive, on-demand, protein synthesis system. In this work, we propose a genetic glucose sensor circuit to direct the downstream synthesis of insulin or glucagon. This circuit can be used for POC treatment of diabetes.

A number of effective metabolic engineering strategies have made use of

the advances in promoter engineering to dynamically regulate the expression of pathway enzymes to optimize product yields [28, 58, 80, 91]. Allosteric transcription factors (aTFs) are a core component of this. An aTF is a protein that binds to the operator sequence of a DNA and prevents transcription. Upon binding to a specific ligand, the aTF undergoes a conformational change that alters its affinity for an operator DNA sequence [167]. Ligand binding leads to the release of the aTF from the operator, allowing the RNA polymerase holoenzyme to effectively bind to the promoter and increase the transcription rate. aTFs have been co-opted for use as gene expression switches [96].

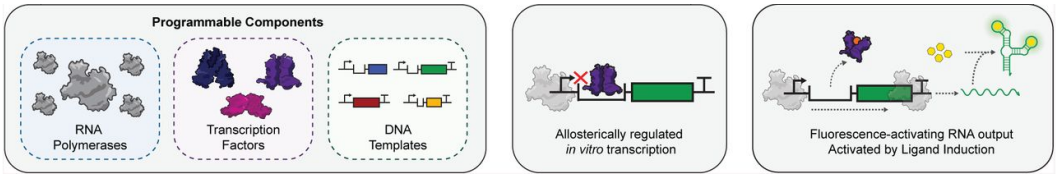


Figure 2.1: In this image, the first box indicates the components involved in an aTF regulated circuit: the RNA polymerases, aTFs, and DNA transcription templates. The second box indicates the circuit in the absence of the ligand. In the presence of the ligand, the aTF will no longer bind to the operator sequence and allows expression of the downstream reporter as shown in the third box. [72]

More recently, aTFs have been increasingly leveraged for use in POC biosensors. CFPS is a favorable platform for implementing biosensors because of its portability, ability to manipulate the reaction environment in real-time, and tolerance to cytotoxic environments. aTFs can be used to 'sense' a specific molecule (the ligand that suppresses it) by altering the gene expression of a downstream reporter protein. By monitoring the signal from the reporter, it would be possible to determine the amount of ligand present in the system. Such biosensors have recently been used for detecting ions, biomarkers, and antibiotics (See Figure 2.1) [72, 153, 182].

The ability to sense glucose concentration in the bloodstream and direct the synthesis of either insulin or glucagon (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 aim to design a biosensing system that has the ability to sense glucose concentration in the bloodstream and direct the synthesis of either insulin or glucagon hormone (or a combination of both). In order to achieve this, we need to construct a transcriptional unit that can effectively respond to variation in the concentration of glucose or its derivatives (such as D-gluconate) that can act as its proxy.

We plan to study the effectiveness of aTFs towards this purpose. An aTF has a DNA binding region at the N-terminal and a ligand-binding region at the C-terminal. Studies on the Gluconate operon transcriptional repressor (GntR) family of *E. coli* and *B. subtilis* have shown that two derivatives of glucose, D-glucono-1,5-lactone and D-gluconate, can bind to GntR protein at the C-terminal and as an anti-repressor [39, 124, 130, 148, 173]. Since our biosensor should be able to sense the presence of glucose in the system we decided to make use of the GntR family in the biosensor.

In this work, we aimed to first construct a promoter (incorporating the operator sequence of GntR) 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 [143]. Next, we added a reporter protein gene downstream of this modified promoter to easily

characterize the promoter dynamics.

Earlier work done by my colleague, Abhinav Adhikari in the VarnerLab made use of a deGFP reporter constructed with an mP70 promoter. This work studied the repression and de-repression performance of GntR in the myTXTL system. However, the D-gluconate was consumed by the metabolism in the myTXTL system resulting in poor de-repression. So in this work, we plan to make use of the PURE cell-free system which lacks the metabolic capability to unintentionally consume D-gluconate. The PURExpress system also makes it easier to purify the protein products. Since all the components of the system (other than the ribosomes) are his-tagged, protein products in the native form (with no His tag) can be purified easily and rapidly through ultrafiltration (to remove the ribosomes) followed by the use of Ni-NTA affinity column (to remove the His-tagged components).

In addition to this, we also changed the reporter protein. We use the Yellow Fluorescent Protein, Venus, instead of deGFP because Venus is a much brighter reporter protein in this system, the time it took to reach maximum growth was similar to different mutated GFP variants [87].

The CFPS reactions to express Venus 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 the PURExpress reconstituted system based on PURE. Taken together, our results show that the GntR protein effectively represses the expression of the Venus reporter protein, and that the addition of D-gluconate is able to inhibit this repression.

2.2 Materials and Methods

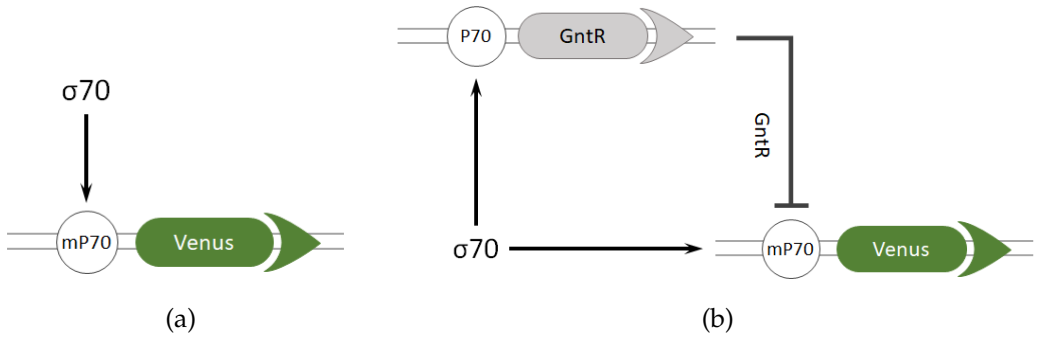
2.2.1 mP70 promoter construction

The first step in constructing the promoter responsive to D-gluconate levels was to include the operator sequence for the E.coli GntR in the promoter region. The sequences of the operator and the repressor were obtained from the literature [124]. The 16 bp *E. coli* GntR operator sequence (ATGTTACCCGTATCAT) was added between the -35 and -10 regions of mutated bacteriophage lambda promoter (P70) developed by Noireaux and coworkers [143]. This modified P70 promoter with added GntR operator site was named mP70. The -10 binding site of the RpoD holoenzyme was altered from GATAAT in the original promoter to TATCAT in the mP70. The gene sequence of the reporter protein, 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′ UTR for the DNA with the mP70 promoter were identical to that of the DNA with the P70 promoter. 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.

2.2.2 Synthetic Circuits Architecture

The three genetic circuits (C1, C2, and C3) used in this study were based upon the bacterial sigma factor regulatory system. Sigma factor 70 (σ70) was the primary driver of each circuit. The E.coli RNA Polymerase Holoenzyme is the core enzyme saturated with σ70 and allows for RNA synthesis from σ70 specific pro-

moters. In C1, σ70 induced Venus (YFP) expression was explored in the absence of additional regulators (Figure 2.2(a)). In C2, σ70 induced the expression of GntR and Venus (Figure 2.2(b)). The GntR protein repressed the mP70 promoter of Venus, thereby down-regulating Venus transcription. In C3, D-gluconate was introduced into the circuit (Figure 2.3). D-gluconate suppresses the binding of GntR to the mP70 promoter. This de-represses the Venus transcription. Studying C1 allows us to estimate parameters governing the interaction of σ70 with the mP70 promoter. Whereas, the C2 allows us to characterize the strength of the transcriptional repression by GntR. Finally, C3 allows us to characterize the de-repression effect D-gluconate has on GntR.



(a) (b)

Figure 2.2: (a) C1 Circuit: σ70 induces Venus expression (b) C2 Circuit: σ70 induces the expression of GntR and Venus. The GntR protein once expressed represses the mP70 promoter of Venus.

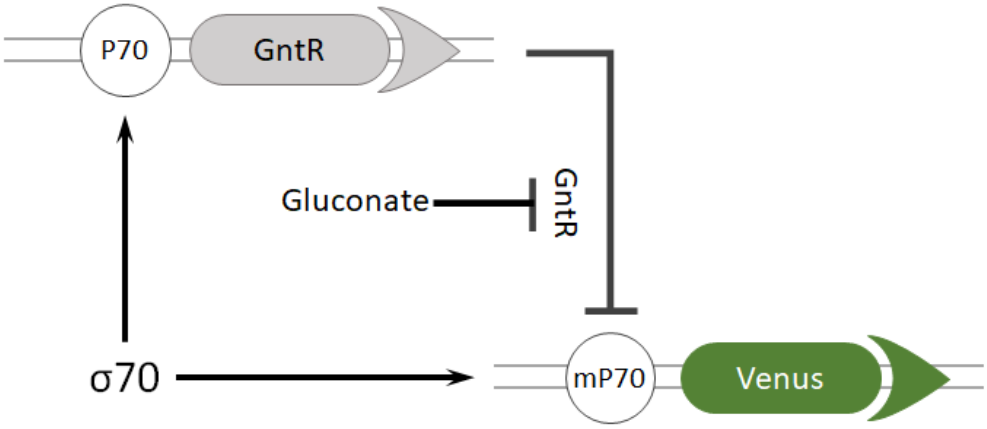


Figure 2.3: C3 circuit: σ70 induces the expression of GntR and Venus. The GntR protein once expressed represses the mP70 promoter of Venus. In the presence of a high concentration of D-Gluconate, the binding of GntR to mP70 gets repressed thereby de-repressing Venus transcription

2.2.3 Cell-Free Protein Synthesis Reactions

The cell-free protein synthesis (CFPS) reactions were carried out using the PURExpress In Vitro Protein Synthesis Kit (New England Biolabs Inc) in 120 μL 384-well plates (Thermofisher NUNC, flat-bottom) in Varioskan Lux plate reader at 37°C. The working volume of all the reactions was 25 μL, composed of the PURE solutions A and B (17.5 μL), RNase inhibitor, Murine (0.5 μL), NEB E.coli RNA Polymerase holoenzyme (2 μL), the linear DNA: Venus (7 nM), GntR (10 nM). The *E. coli* RNA Polymerase, Holoenzyme is the core enzyme saturated with sigma factor 70 and initiates RNA synthesis from sigma 70 specific promoters. 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. For the control reactions without GntR, an equal volume of nuclease-free water was used in its stead. For the de-repression reactions,

D-gluconate (10 mM) was added while maintaining the same total volume.

2.2.4 mRNA Quantification

Following each CFPS run, the total RNA was extracted from 5 μL 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 DNA (Venus and GntR) were carried out for 2 hr. 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.

2.2.5 Protein Quantification

Venus (Yellow Fluorescent Protein) fluorescence was measured using the Varioskan Lux plate reader at 513 nm (excitation) and 531 nm (emission). The fluorescence was measured with 25 μL for each of the mixtures. For all measurements, at least three biological replicates were performed. A protein standard

curve was used to determine the absolute protein concentrations for each of the samples. The standard curve was prepared by using Venus standards of known Relative Fluorescence Units (RFU). An SDS-PAGE of these standards along with a Protein ladder of samples with known concentrations was performed at 300V until the dye fronts reached the bottom of the gel. The SDS-PAGE was performed in a Mini-PROTEAN Tetra Cell (Bio-Rad Laboratories) using Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 15 well, 15 µl gels (Bio-Rad Laboratories). This was followed by staining the gel with Bio-Safe™ Coomassie Stain (Bio-Rad Laboratories). The stained gel was visualized using the Chemi-Doc MP and then the concentrations were measured using Image Lab software.

2.2.6 Estimation of model parameters

To model the circuit, the mRNA and protein balances for mRNA m_i and protein p_i are described using the following balance equations

$$\dot{m}_i = r_{X,i} u_i - \theta_{m,i} m_i \tag{2.1}$$

$$\dot{p}_i = r_{L,i} w_i - \theta_{p,i} p_i \tag{2.2}$$

where r_X and r_L are the transcription and translation rates and θ_m and θ_p are the respective degradation rates. Here, w_i and u_i are the control functions modeled as:

$$\dot{w}_i = -\left(\frac{ln(2)}{\tau_{L,1/2}}\right) w_i \tag{2.3}$$

$$u_i = \frac{\sum_{j \in \chi} W_j f_j(...)}{\sum_{i \in C_i} W_i f_i(...)} \tag{2.4}$$

W_i denotes the weight of configuration i and f_i is the binding function modeled as a hill function. The rate of transcription $r_{X,j} u_j$ was modeled as the product of a

kinetic limit, $r_{X,j}$, and a control term $u_j \in [0, 1]$. The kinetic limit of transcription was derived from elementary reactions leading to the formation of m_j, as previously done by McClure [101]. The kinetic limit of transcription was formulated as:

$$r_{X,j} = V_{X,j}^{max}\left(\frac{\mathcal{G}j}{\tau X, jK_{X,j} + \left(\tau_{X,j} + 1\right)\mathcal{G}_j}\right) \tag{2.5}$$

where the initiation time constant $\tau_{X,j}$, and saturation constant $K_{X,j}$ were experimentally measured by McClure [101]. The maximum transcription rate, $V_{X,j}^{max}$, was given by:

$$V_{X,j}^{max} = \left[R_{X,T}\left(\frac{\dot{v}X}{lG, j}\right)\right] \tag{2.6}$$

where $R_{X,T}$ is the RNA polymerase concentration (nM), $\dot{v}X$ is the RNA polymerase elongation rate (nt/h) and lG, j is the length of gene j in nucleotides (nt). By analogy, the kinetic limit of translation was formulated as:

$$r_{L,j} = V_{L,j}^{max}\left(\frac{x_{mRNA}}{\tau_{L,j}K_{L,j} + \left(\tau_{L,j} + 1\right)x_{mRNA}}\right) \tag{2.7}$$

where x_{mRNA} is the mRNA concentration and $V_{L,j}^{max}$ is the maximum translation rate which was formulated as:

$$V_{L,j}^{max} = \left[K_P R_{L,T}\left(\frac{\dot{v}L}{lP, j}\right)\right] \tag{2.8}$$

The model previously used by Adhikari [2] was extremely effective in capturing the expression dynamics of mRNA and protein in synthetic cell-free circuits, but it suffered from one major drawback. The parameters related to the Gibbs free energy of binding between RNAP and other genes in the systems were overestimated. This was primarily due to the lack of literature that provides the constraints for ΔG. Since the Gibbs free energy of binding drives the weight of a particular configuration, any overestimation will lead to a high background expression. This will affect the transcription rate which will thereby affect the

translation rate and the protein expression. Therefore, to address this issue, we revised the model by changing the way the control function for transcription is formulated. A similar modeling approach as previously mentioned was adopted but the control functions were formulated as -

$$u_i = \frac{\sum_{j \in \chi} W_j f_j(...)}{\sum_{i \in C_i} W_i f_i(...)} \tag{2.9}$$

where W was formulated as-

$$W_i = exp\left(\frac{-\Delta E_i}{k_b T}\right) \tag{2.10}$$

This had a two-fold advantage. Firstly, there were literature estimates available for difference in binding energy (ΔE) from Bintu and coworkers [12], in reference units, as multiples of Boltzmann constant (k_b) and temperature (T). Moreover, this formulation was much closer to the statistical mechanics approach we have undertaken to represent the configurations.

2.3 Results and Discussion

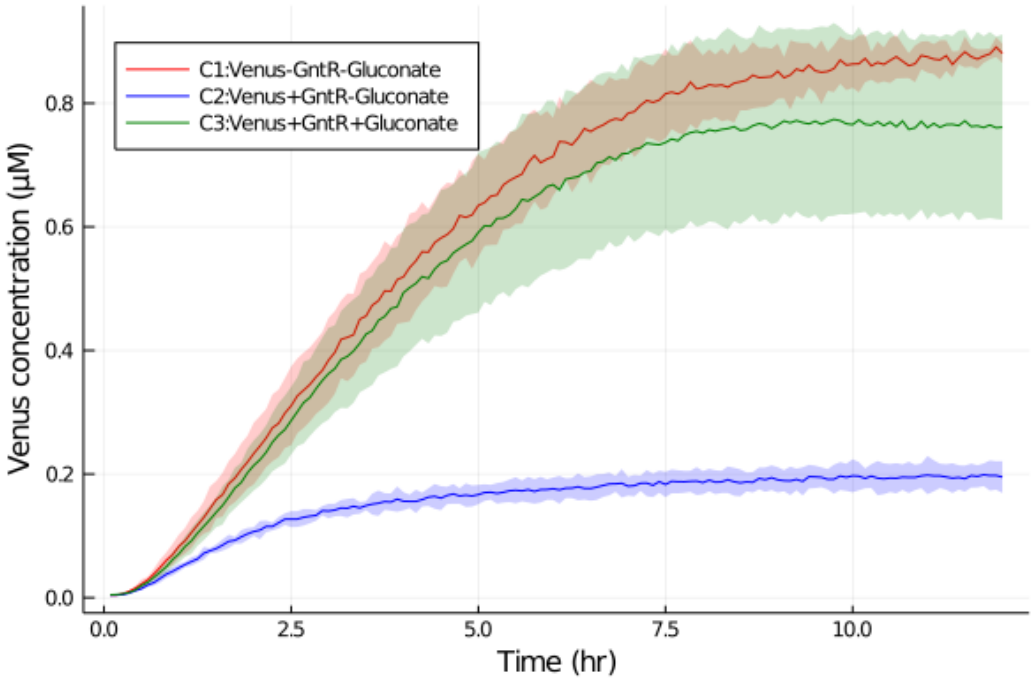
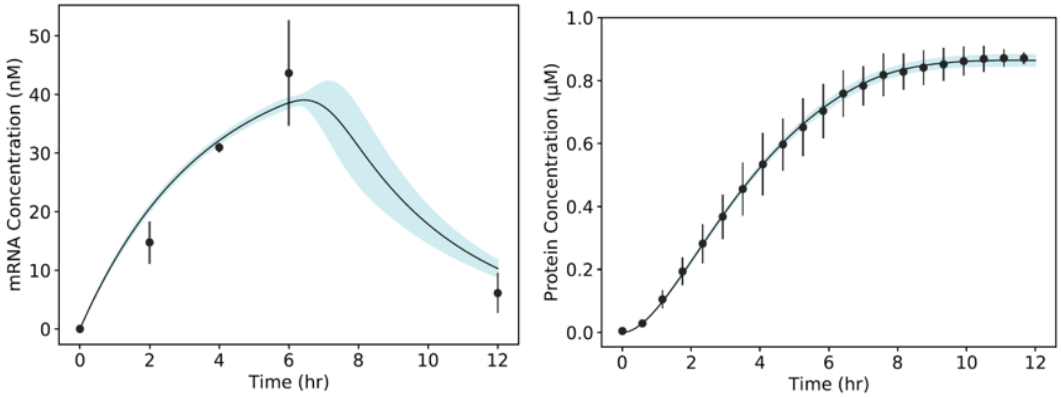


Figure 2.4: Venus protein expression without GntR (C1 - Red), Venus repression by GntR (C2 - Blue), and de-repression by D-gluconate (C3 - Green). The shaded regions represent one standard deviation of the experimental measurements

We first tested the performance of linear DNA with the mP70-Venus construct (7 nM) in the PURExpress system (C1 circuit) in order to obtain a standard for comparison for the repression and de-repression characteristics. In order to characterize the repression by Gntr, we made use of the C2 circuit. We expressed two linear DNA fragments: *E. coli* GntR (10 nM) and mP70-Venus (7 nM). For the control case, the GntR DNA was excluded from the reaction mixture. While the Venus concentration in the control case steadily increased, it reached a steady plateau in approximately 4 hours in the +GntR reaction (well before the -GntR control plateaus) and had a significant reduction in Venus expression. This showed that GntR was indeed repressing Venus expression.

In order to test the de-repression characteristic of GntR, we added 10mM of D-gluconate (approximately 9 times the value of its K_D established in literature [23]) along with GntR and Venus (C3 circuit). Although the Venus concentration was not as high as in the case of the C1 circuit, it was appreciably higher compared to when D-gluconate was not added (C2 circuit).

2.3.1 Parameter Estimation

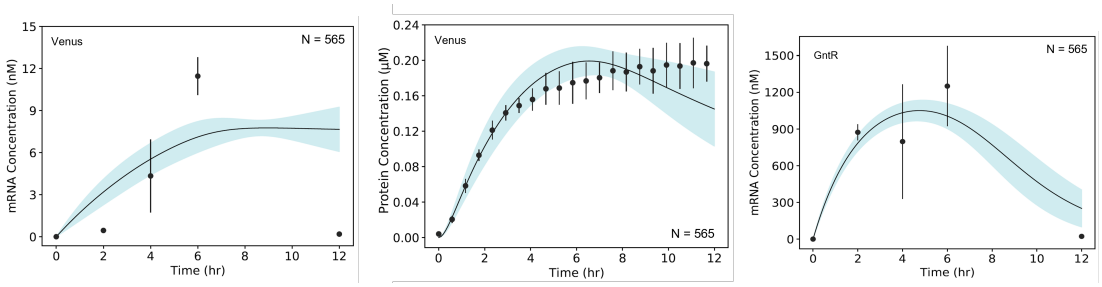


(a) (b)

Figure 2.5: Model simulations versus experimental measurements for σ70 induced Venus expression for the C1 circuit. (a) Simulated and measured Venus mRNA concentration versus time. (b) Simulated and measured Venus protein concentration versus time. The black points indicate the experimental values while the black curve is the predicted curve. The shaded region denotes the 95% confidence interval of the predicted curve. The error bars denote one standard deviation of the experimental data

Parameter estimation for the C1 circuit was set up as a multi-objective optimization problem with two objectives. One objective was the minimization of the error between experimental and simulated values of Venus mRNA, and the other for the corresponding Venus protein measurements. The algorithm was run for 20 generations and produced an ensemble of 187 parameter sets, of the

11 unknown model parameters. The criterion for a parameter ensemble selection was a Pareto rank of 2 or below. The parameters generated for the C1 circuit were able to create a good fit with the experimental data (Figure 2.5).



(a) (b) (c)

Figure 2.6: Model simulations versus experimental measurements for the C2 circuit. (a) Simulated and measured Venus mRNA concentration versus time. (b) Simulated and measured Venus protein concentration versus time. (c) Simulated and measured Gntr mRNA concentration versus time. The black points indicate the experimental values while the black curve is the predicted curve. The shaded region denotes the 95% confidence interval of the predicted curve. The error bars denote one standard deviation of the experimental data

In the case of the C2 circuit, the multi-objective optimization problem had three objectives: Minimization of error between experimental and simulated values of (i)Venus protein (ii) Venus mRNA, and (iii) GntR mRNA. The algorithm was run for 10 generations to estimate 22 unknown model parameters and estimated an ensemble of 565 parameter sets that had a Pareto rank of 2 or below. Although less effectively than for the C1 circuit, the model was able to capture the mRNA concentration profile for Venus fairly well. The model accurately captured the mRNA concentration profile of GntR and protein expression of Venus. Taken together, the model is able to simulate the repression of mP70-Venus in the presence of GntR in a PURE system.

2.3.2 Future Directions

The D-gluconate levels at the different time points can be measured by Liquid Chromatography/Mass Spectrometry (LCMS) using our lab's published derivatization protocol [180]. D-gluconate acts as a dynamic second messenger for D-glucose concentration in myTXTL cell-free extracts. We used the gluconate-responsive transcription factor 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 capability of the proposed circuit. We also need to construct a dose-response curve to study the de-repression of GntR in the circuit at different concentrations of D-gluconate. This would be necessary to better optimize the switch behavior of the circuit.

Secondly, in order to fully characterize repression and de-repression, we plan to use different concentrations of the DNA construct with the mP70 promoter and the GntR protein to perform Electrophoretic Mobility Shift Assays (EMSA) in order to determine the binding parameters between Gntr protein and mP70. We also plan to do end point measurements of the reporter fluorescence for these concentrations. We also plan to implement real-time measurement of mRNA to get more mRNA measurements making it easier to develop accurate models [115].

Thirdly, we also plan to use different GntR constructs derived from *P. aeruginosa* and *B. subtilis* and test their effectiveness in repressing reporter fluorescence. The *P. aeruginosa* construct was shown to be partly responsive to D-glucose, eliminating the need for a proxy, D-gluconate [23]. We are also in the process of testing the effectiveness of a circuit which uses promoters recognized

by T7 polymerase instead of the *E. coli* RNAP. The T7 polymerase provides a two-fold advantage over the *E. coli* RNAP: It synthesizes RNA at a much higher rate [161]; It is inherently present in the PURE system and doesn't need to be added separately [142].

Fourthly, we plan to develop a mathematical model for the C3 circuit to take into effect the presence of D-gluconate in the system to effectively characterize transcription and translation behavior. We would have to estimate the model parameters using multi-objective optimization similar to how they were determined for C1 and C2 circuits. Such a model would also help guide our future experiments based on this system.

Lastly, we also plan to further engineer the GntR protein (and the promoter region) to enable more efficient repression and derepression properties. Mutations in the C-terminal region of GntR have been shown to affect its ability to bind to D-gluconate while mutations in the N-terminal region affect operator binding properties [196, 197]. On the basis of these details, we plan to generate a library of GntR protein with random mutations in the C-terminal and N-terminal regions via error-prone PCR in order to discover GntR mutants that have good operator binding strength, while still being effectively antagonized by D-gluconate.

Once we have accomplished these, we can test the circuit with fluorescent labeled insulin and glucagon. The ultimate aim is to make use of this transcriptional unit to construct a viable glucose biosensor (Figure 2.7) that is capable of producing insulin at high glucose (gluconate) levels and glucagon at high glucose (gluconate) levels.

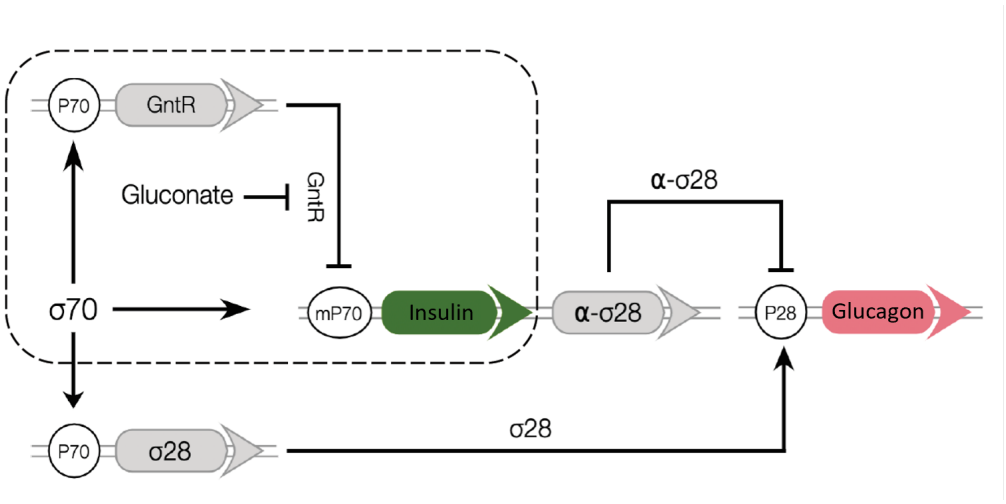


Figure 2.7: Schematic of a glucose biosensor: σ70 induces the expression of GntR, σ_{28}, insulin and α-σ_{28} which is downstream to insulin. σ_{28} induces expression of glucagon while α-σ_{28} inhibits the P28 promoter thereby inhibiting the expression of glucagon. When the gluconate levels are high, insulin will be expressed along with σ28. At low gluconate levels, only glucagon will be expressed

CHAPTER 3

I(NSP)ECTION OF THE SARS-COV-2 NSP1'S HOST-TRANSLATION EFFECT IN A CELL-FREE SYSTEM

3.1 Introduction

The SARS-CoV-2 virus which is the cause of the unprecedented Covid-19 pandemic belongs to the class of coronaviruses. Coronaviruses are enveloped viruses with a positive-sense single-stranded RNA genome. This mRNA contains a 5′ cap structure along with a 3′ poly (A) tail allowing it to serve as the genetic material as well as the messenger RNA which translates to the proteins [36,183].

The SARS-CoV-2 has a genome of ~30 kb. It possesses 14 Open Reading Frames (ORFs) encoding for a total of 27 known proteins [75,78,131]. The overlapping ORFs ORF1a and ORF1ab located at the 5′ terminal of the genome encode the pp1a and pp1ab polyproteins respectively. The translation of ORF1ab is mediated by a -1 ribosomal frameshift at the overlap of these two ORFs [37,183]. The other ORFs are located at the 3′ terminal and encode for the 4 structural proteins (Spike protein (S), Small Envelope protein (E), Matrix protein (M), Nucleocapsid (N)) and 8 accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, ORF14) [183,184,191].

The pp1a and pp1ab polyproteins together contain 15 functional non-structural proteins (nsp) (nsp11 is a 13 amino acid cleavage product that isn't known to have any function [109,159]). These nsps are released from pp1a and pp1ab upon proteolytic cleavage by nsp3 and nsp5 [37,183].

The non-structural protein 1 (nsp1), also known as the host shutoff factor, is encoded by the gene closest to the 5′ end of the viral genome and is one of the first proteins to be expressed after the virus infects a host cell [138,198]. It is one of the most important nsps that is necessary for the propagation of the virus. It inhibits host gene expression and helps the virus suppress the host immune response [110,159,169,183].

In other coronaviruses like the SARS-CoV and MERS-CoV, the nsp1 inhibits host translation by forming a complex with the 40S ribosomal unit which renders this inactive for host translation [66,108,110,111,138,165,169,198]. Further, this also causes a selective endonucleolytic cleavage in the 5′ untranslated region (UTR) of the host mRNA [61,108,169]. This promotes degradation of the host mRNA while the viral mRNA remains intact.

The SARS-CoV-2 nsp1 has ~84% amino acid sequence similarity with its SARS-CoV homolog [134, 150, 159, 169]. Such high sequence similarity suggests that they share common properties and functions. The SARS-CoV-2 nsp1 binds to the human 40S ribosomal subunit to shut off host translation. The protein inserts its C-terminal domain into the mRNA entry channel of the ribosome (at the 18S ribosomal RNA), where it interferes with the host mRNA binding [11, 83, 138, 150, 169]. A stem-loop structure of SARS-CoV-2 mRNA 5′UTR has been observed that is recognized by nsp1. This recognition enables selective translation of the viral mRNA while still inhibiting host translation [134,138,150,169]. Thus, nsp1 also helps to boost viral protein production by hijacking more of the cellular machinery for translation of viral proteins [11]. This recognition also protects the viral mRNA from the possible mRNA degradation activity of the nsp1 [61,73,165].

Type I interferon (IFN) induction and signaling represents one of the major innate antiviral defense mechanisms and many viruses have developed strategies to evade this immune response [54, 110]. The expression of these IFNs triggers innate antiviral immune responses which help to suppress the viral replication. Coronavirus infections are sensed by RIG-I (Retinoic acid–Inducible Gene I)-like receptors (RLRs), which activate this IFN-based defense mechanism [129, 150, 154, 169]. nsp1's translation inhibition functions interfere with the production of these proteins by shutting down the host translation machinery. Thus by affecting the translation of antiviral defense factors such as IFN-β or RIG-I, nsp1 effectively weakens the innate immune response [166, 169].

Its central role in the suppression of host translation and the immune response makes nsp1 a viable target for the development of anti-viral drugs and further studies. In this work, we aimed to produce the SARS-CoV-2 nsp1 using E.coli-based CFPS. Secondly, we wanted to test the potential of nsp1 to inhibit host translation in the same cell-free system (See Figure 3.1). And then mathematically model the effect of nsp1 on the host translation if it turned out to demonstrate statistically significant inhibition.

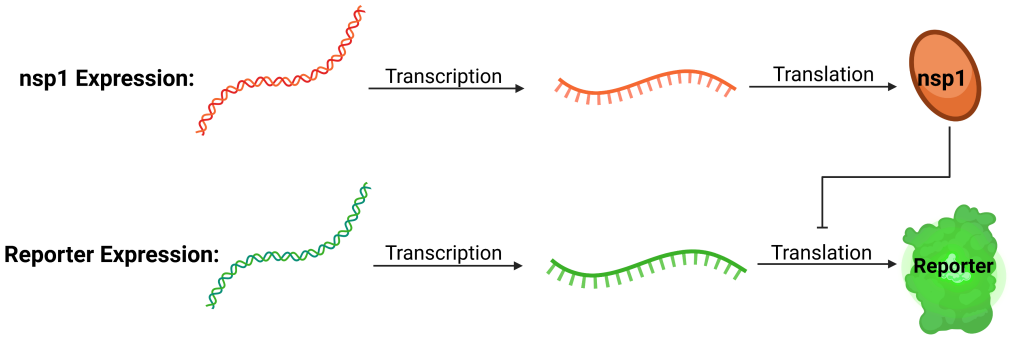


Figure 3.1: Circuit representing coexpression of nsp1 along with a reporter protein. The nsp1 binds to the ribosomes and selectively prevents translation of the reporter protein. This circuit doesn't include the potential mRNA degradation activity of nsp1 [61, 73, 165]. (Created with BioRender.com)

To this purpose, we designed a DNA construct that was ”reverse-translated” from the sequence of the protein [200]. We used a P70 promoter sequence, along with its 5′ UTR, as it is optimized to maximize expression in the *E. coli*-based myTXTL Cell-free system [42]. We also carried out an assay to test the ability of nsp1 to inhibit translation of dual emission Green Fluorescent Protein (deGFP) (which would take the place of a host protein) in myTXTL extract.

3.2 Materials and Methods

3.2.1 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 12.8 μL, composed of the Sigma 70 Master Mix (9 μL), GamS nuclease inhibitor (0.8 μL), and the nsp1 linear DNA fragments (3 μL - 9.38nM). 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. The CFPS reactions were incubated at 29°C. The CFPS runs were carried out in triplicate for 2hr, 4hr, 8hr, and 16hr time points. Following each CFPS run, the samples were stored at -80°C.

3.2.2 mRNA Quantification

Following each CFPS run, the total RNA was extracted from 5 μL 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: CFPS reaction for 5 nM of nsp1 DNA was carried out for 2 hr. 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.

3.2.3 Protein Quantification

Following each CFPS run, 2 μL of each sample was extracted and added to 4 μL of water. And an equal amount of 2x Laemmli sample buffer (with 2-mercaptoethanol as a reducing agent) was added to these. An SDS-PAGE of 10μL of each of these diluted samples along with a Protein ladder of samples with known concentrations was performed at 300V until the dye fronts reached the bottom of the gel. This was followed by staining the gel with Bio-Safe™ Coomassie Stain (Bio-Rad Laboratories). The SDS-PAGE was performed in a Mini-PROTEAN Tetra Cell (Bio-Rad Laboratories) using Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 15 well, 15 µl gels (Bio-Rad Laboratories). The stained gel was visualized using the ChemiDoc MP and then the concentrations were measured using Image Lab software.

3.2.4 Coexpression of deGFP with nsp1

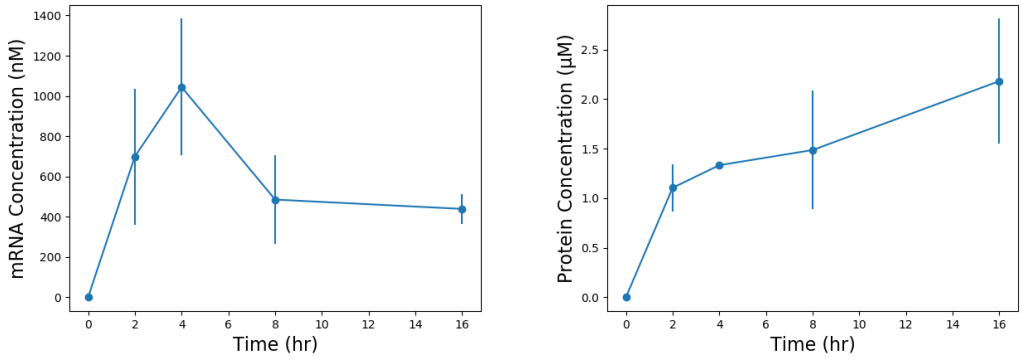
The cell-free protein synthesis (CFPS) reactions were carried out using the myTXTL Sigma 70 Master Mix (Arbor Biosciences) in 120 μL 384-well plates (Thermofisher NUNC, flat-bottom) in Varioskan Lux plate reader at 29°C. The working volume of all the reactions was 14.1 μL, composed of the myTXTl extract (9 μL), GamS nuclease inhibitor (0.6 μL), the linear DNA: deGFP (3 μL - 4.26 nM), nsp1 (1.5 μL - 4.26 nM). Two sets of control reactions were also run. In one, instead of nsp1, an equal volume of water was used. And in the other, equal volume and concentration of nsp8 linear DNA construct was used instead of nsp1. nsp8 was chosen as a control as it has a size similar to that of nsp1 but doesn't inhibit translation unlike nsp1 (though it could affect protein trafficking to the cell membrane) [11]. The Cell-free reactions were carried out in duplicates for 16hr. deGFP fluorescence was measured using the Varioskan Lux plate reader at 488 nm (excitation) and 535 nm (emission) at 5 min time intervals. In order to confirm the expression of these proteins, following the CFPS run, 2 μL of each sample was extracted and diluted with 3 μL of water. Following this 5 μL of 2x Laemmli sample buffer was added to these and an SDS-Page of these samples was performed at 200V until the dye fronts reached the bottom of the gel. This was followed by Coomassie staining of the gel and visualization using the Image Lab software.

3.2.5 Qualitative testing of nsp1 binding with E.coli Ribosomes

We ran a CFPS reaction of His_6-tagged nsp1 in myTXTL. The working volume of the reaction was 12.5 μL, composed of the Sigma 70 Master Mix (9 μL), GamS

nuclease inhibitor (0.5 μL), and the nsp1 linear DNA fragments (3 μL - 9.6nM). The CFPS reactions were incubated at 29°C for 12hr. Then we purified the His_6-tagged nsp1 using Ni-NTA affinity column. Meanwhile, we separated the E.coli Ribosomes from the PURExpress In Vitro Protein Synthesis Kit (New England Biolabs Inc). Then these ribosomes were added to half the volume of the purified nsp1 proteins and allowed to mix well at 29°C for 15 minutes. Then this sample was loaded into a gel along with pure nsp1 protein samples and a Protein ladder for an SDS-PAGE run. The ribosomes and its subunits are expected to be retained at the top while the nsp1 should be able to pass through. After the SDS-PAGE run, the gel was stained with Bio-Safe™ Coomassie Stain (Bio-Rad Laboratories) and the stained gel was visualized using the ChemiDoc MP. The concentration of nsp1 in each well was measured using Image Lab software (and adjusted for the dilution of nsp1 in the channel with the ribosomes).

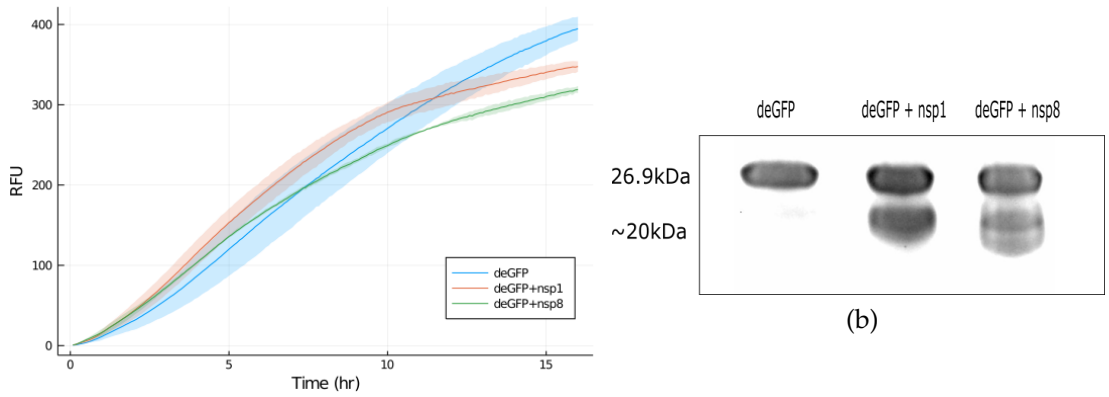
3.3 Results and Discussion



(a) (b)

Figure 3.2: Gene expression of nsp1. The points denote the mean, while the error bars denote one standard deviation calculated from three replicates

We were able to successfully synthesize SARS-CoV-2 nsp1 in myTXTL extract. With a starting concentration of 9.38 nM, we were able to obtain up to an average of 2.18 μM of nsp1 protein after a CFPS run of 16hr. The mRNA concentration peaked at about 1010 nM at the 4hr measurement point and continued to decline after that (Figure 3.2).



(a) (b)

Figure 3.3: Coexpression of deGFP with nsp1: (a) RFU measurements of deGFP when coexpressed with nsp1 (Red). The controls: expression of deGFP alone (Blue), expression of deGFP along with nsp8 (Green). The shaded regions indicate one standard deviation (b) SDS-PAGE Analysis of these samples to confirm the production of all the proteins. The gels have been Coomassie-stained

From Figure 3.3, we can observe that although coexpressing nsp1 along with deGFP does reduce the synthesis of deGFP, there isn't a very significant difference from just expressing deGFP by itself. In addition to that, coexpressing deGFP with a control gene coding for the nsp8 protein with a molecular weight (~ 22.8kDa) just slightly higher than nsp1 (~ 19.77kDa) [200] also reduces the production of deGFP. This is despite the fact that the nsp8 doesn't inhibit host translation [11]. This decrease in the production of deGFP in presence of nsp1 or nsp8 can be better explained by a competition for the metabolic resources caused by the introduction of their genes.

This also goes to prove that nsp1 doesn't affect the prokaryotic ribosomes. Unlike the larger mammalian 80S ribosome made up of 60S and 40S subunits, the prokaryotic 70S ribosome is made up of 50S and 30S subunits. The nsp1 which binds to the mRNA entry channel of the 40S subunit doesn't bind to the corresponding region in the prokaryotic 30S subunit. Therefore, it doesn't actually inhibit translation in the E.coli based myTXTL system.

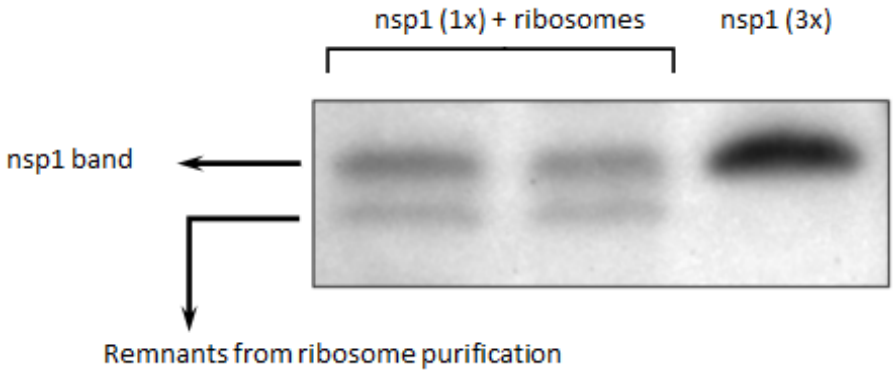


Figure 3.4: Rudimentary nsp1-70S Ribosome binding assay: The third channel contains pure nsp1. The first and second channels contain nsp1 mixed with two times its volume of purified 70S ribosomes solution. In other words, the concentration of nsp1 in the third channel is thrice the concentration of nsp1 in the first two. Measurement of the fluorescence of the nsp1 bands show that the normalized concentration of nsp1 in all wells are approximately the same indicating that there is negligible binding of nsp1 with these ribosomes

We used a rudimentary binding assay to test the binding of nsp1 with E.coli ribosomes. Ribosomes treated with nsp1 proteins were loaded on two channels of a gel alongside a different channel containing pure nsp1. The channels are expected to retain the ribosomes, their subunits, and any nsp1 bound to it at the top but allow the free unbound nsp1 to pass through [4]. The results of the SDS page show a negligible difference between the channel with just nsp1 and the channels with nsp1 and the ribosome. The channel with only nsp1 had 116.4 ng/μL of nsp1 while the ones where it was treated with ribosome had

an average of 106 ng/μL of nsp1 (See Figure 3.4). If nsp1 had bound to the ribosome, we would have seen a significantly reduced concentration.

These results combined confirm that nsp1 doesn't bind/shows negligible binding to E.coli ribosomes. Conversely, this does prove that we can still synthesize nsp1 using myTXTL cell-free extracts and that this synthesis can be modeled.

3.3.1 Future Directions

There are several directions for this work that can be pursued. First, in order to better characterize this system, we can use the obtained data to fit the parameters for nsp1 gene expression (transcription and translation behavior) in myTXTL using the Pareto Optimal Ensemble Technique. Second, in order to better determine the binding between E.coli ribosomes and nsp1, we can make use of better assays such as Sucrose gradient centrifugation can be used. Lastly, we can repeat these experiments with a mammalian cell-free extract. The nsp1 should be able to bind to the 40S ribosomal subunits of the 80S mammalian ribosomes present in these.

We could also analyze nsp1's binding to ribosomes in cell-free systems based on different mammalian cells such as Rabbit reticulocytes, CHO cells, and HeLa cells [46]. It would also be interesting to study nsp1's inhibition effects in cell-free extracts prepared from human cell lines of different systems/organs. Cells from human lungs are primarily affected by the SARS-CoV-2 virus infection, so human lung cell lines should theoretically provide the best environment for the growth of the virus [136, 141, 172]. We could compare lung-based cell-free

extracts from Calu-3 cells, A549 cells, or HF10B4 cells [103,188], with other human cell-free extracts like HeLa cells, HEK 293 cells, and K562 cells [16,187,199], to observe if there are any noticeable differences between nsp1's effect in these systems.

With these cell-free systems, we can also study the host translation inhibition of nsp1 and create a mathematical model for the system. Since the 5'UTR of the viral genome helps it bypass nsp1's translation inhibition and is necessary for viral replication, the nsp1 gene can be designed with this 5'UTR. The nsp1 can be used as a global translation inhibitor to terminate the production of proteins in a suitable cell-free system.

CHAPTER 4

MODELING THE EFFECT OF THE RNA-DEPENDENT RNA POLYMERASE IN VITRO

4.1 Introduction

The SARS-CoV-2 is a coronavirus and like all coronaviruses, it is a positive-sense single-stranded RNA virus ((+)ssRNA). In other words, its genome consists of a single strand of positive-sense RNA (which goes in the 5′ to 3′ direction) [36, 169,183,184]. This RNA genome serves as the template for both translation and RNA replication.

Coronaviruses use an RNA-dependent RNA polymerase (RdRp) for the replication of their genome. The SARS-CoV-2 RdRp forms a complex made of multiple non-structural proteins (nsps). The core of the RdRp is the nsp12 catalytic subunit [5]. The nsp12 contains an N-terminal nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain, an interface domain, and a C-terminal RdRp domain [57, 86, 125]. The RdRp domain consists of fingers, palm, and thumb subdomains [41, 178]. However, nsp12 has little processivity by itself and in order to function properly, it requires the accessory subunits, nsp7 and nsp8 [41,45,79,158,195]. The nsp7 and nsp8 subunits bind to the thumb, and an additional copy of nsp8 binds to the fingers domain [41,79]. The nsp7-nsp8 heterodimer is believed to be capable of de novo initiation and has been proposed to function as an RNA primase to synthesize small primers that are further extended by the nsp12 [13,62,79,168].

The RdRp also coordinates with many accessory factors such as the nsp13

and the nsp14 in order to carry out its functions [22, 98, 152, 194]. The nsp13 is a helicase that is essential for unwinding the RNA complex after replication in a 5′ ⟶ 3′ direction [63, 64, 85, 140]. Besides this, nsp13 also displays 5′-triphosphatase activity, which suggests a possibility that it is involved in mRNA capping [22,63,64]. The nsp14 which possesses an N-terminal exonuclease (ExoN) domain, together with its co-factor nsp10, forms plays an important role as a proofreading machinery during viral replication [29,105,151,152].

Once the virus enters the host cells, its positive-sense RNA (+mRNA) is able to act as a template immediately without having to undergo any modification. This allows the RdRp to use it to synthesize complementary negative-sense RNA (-mRNA), immediately. The initiation of synthesis occurs at the 3′ end of the template and proceeds in the 5′⟶3′ direction [178]. This complementary strand that is synthesized is then, itself, able to act as a template for the production of new positive-stranded viral genomes. Meanwhile, the positive-sense RNA is also translated using the cellular machinery. These newly synthesized proteins and viral genome are further packaged and released from the cell ready to infect more host cells. The -mRNA does not translate for proteins (or translates to garbage proteins) [89,123,128] (See Figure 4.1).

The key role that RdRp plays in viral replication makes it a promising antiviral drug target [202]. Drugs such as Favipiravir, Remdesivir, and Galidesivir are being developed to treat viral infection by inhibiting the effects of RdRp [3, 81, 112]. For instance, Remdesivir is a nucleoside analog that gets incorporated into RNA when the viral genome is being replicated. This creates a translocation barrier that the RdRp encounters after the addition of three more nucleotides following remdesivir causing the RdRp to stall [45, 81].

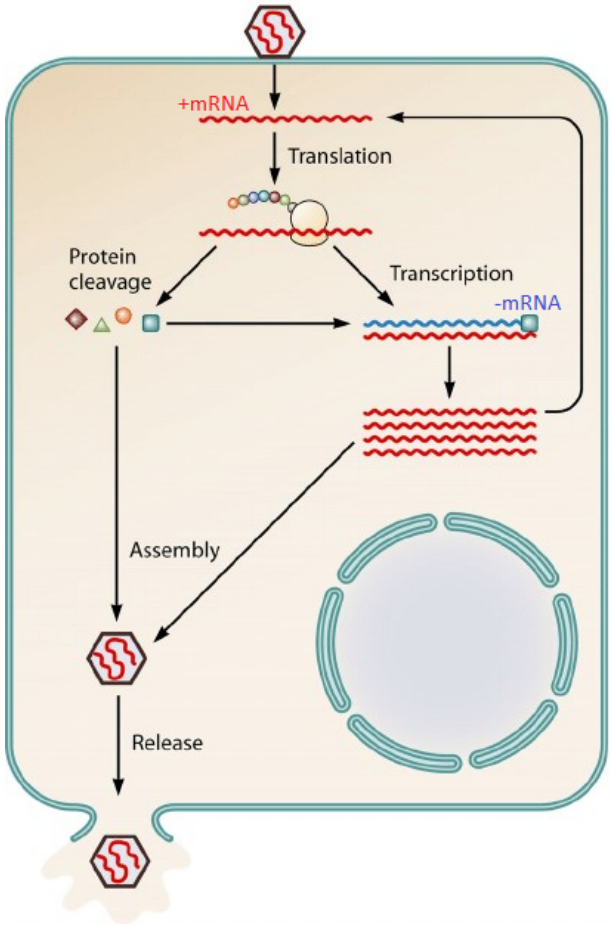


Figure 4.1: Replication strategy of a (+)ssRNA virus: The viral genome is translated using cellular machinery. This genome also acts as a template for the production of complementary -mRNA. This complementary strand then acts as a template for the production of a new viral genome (+mRNA). (Image adapted from [89])

The RdRp of some viruses have been studied to require the presence of certain sequences in the RNA analogous to the DNA promoter-specific transcription by DNA-dependent RNA polymerases [147]. However, in the case of the SARS-CoV-2, although nsp12 can distinguish RNA from DNA, no specific interactions have been observed between base pairs of RNA and residues from nsp12, which suggests that RdRp binding may be RNA sequence-independent [71, 195]. In the event that the activity of the SARS-CoV-2 RdRp is sequence-

dependent, then the 5'UTR and 3'UTR of the genome must be highly conserved in order for the RdRp to produce the complementary RNA strand from the viral genome and also produce the viral genome from this complementary strand [95]. This also indicates that the RdRp can be harnessed to boost the production of any desired proteins.

In this work, we aimed to simulate the effect of RdRp in a cell-free system. To this purpose, we introduce the RdRp into a simple genetic circuit in a myTXTL based cell-free system and simulate its effect to amplify mRNA concentration and thereby improving protein production. We also attempted to produce the subunits of the RdRp complex using myTXTL in order to study its effects in primer extension and expression of a reporter protein.

4.2 Materials and Methods

4.2.1 Model Derivation

In a typical gene expression system, the overall flow of information starts from the coding region of DNA, which is read by the RNA polymerase, to produce a messenger RNA molecule (Transcription - TX) which is then read by a ribosome to produce a protein (Translation - TL). Along the way, both the mRNA and the protein can degrade (at different rates) [177]

If we consider a system with N genes general, the gene expression of each gene G_i is described by two differential equations: one for the rate of change of mRNA 'i' concentration ($\dot{m}_i$), and one for the rate of change of protein 'i'

concentration ($\dot{p}_i$):

$$\dot{m}_i = r_{X,i}\bar{u}_i - (\mu + \theta_{m,i})m_i + \lambda_i \tag{4.1}$$

$$\dot{p}_i = r_{L,i}\bar{w}_i - (\mu + \theta_{p,i})p_i \tag{4.2}$$

The term $r_{X,i}\bar{u}_i$ (nM/hr) denotes the regulated production rate of mRNA 'i' by transcription. It is the product of the kinetic limit of transcription $r_{X,i}$ (nM/hr) and a transcriptional control term $\bar{u}_i$ Similarly, the term $r_{L,i}\bar{w}_i$ (nM/hr) denotes the regulated production rate of protein 'i' translation. It is the product of the kinetic limit of translation $r_{L,i}$ (nM/hr) and a translational control term $\bar{w}_i$. The term λ_i denotes the unregulated rate of transcription (the leak for gene G_i). The $\bar{u}$ and $\bar{w}$ terms describe the control logic of the cell for transcription and translation respectively. These terms are dimensionless, and bounded between 0 and 1. The term μ denotes the dilution term (hr^{-1}), while the terms $\theta_{m,i}$ and $\theta_{p,i}$ the first-order mRNA degradation rate constant (hr^{-1}) and protein degradation rate constant (hr^{-1}) respectively. [2, 177]

Since this reaction is in a cell-free system, the dilution (μ) term can be removed and we assume no leak (λ_i). So these can be removed to modify the equations to :

$$\dot{m}_i = r_{X,i}\bar{u}_i - \theta_{m,i}m_i \tag{4.3}$$

$$\dot{p}_i = r_{L,i}\bar{w}_i - \theta_{p,i}p_i \tag{4.4}$$

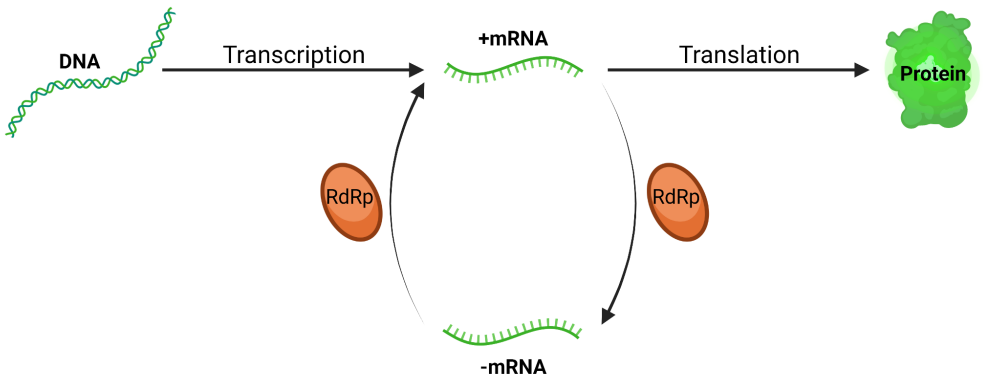


Figure 4.2: Introduction of RdRp into a genetic circuit: RdRp uses the +mRNA as a template for the production of a complementary -mRNA and uses the -mRNA as a template for the production of +mRNA. Meanwhile, the +mRNA is also translated to synthesize proteins (Created with BioRender.com)

In order to study the effect of the RdRp, we modified the mRNA rate equation to include the transcription by RdRp. In this case, we assumed the RdRp to be a complex of nsp7, nsp8, nsp12, and nsp13. The changes brought by RdRp can be applied for both +mRNA and -mRNA :

$$\dot{m}_i^+ = r_{X,i}\bar{u}_i - \theta_{m,i}m_i^+ + f(m_i^-, R_d) \quad (4.5) \qquad \dot{m}_i^- = -\theta_{m,i}m_i^- + f(m_i^+, R_d) \quad (4.6)$$

For the +mRNA, this additional function $f(m_i^-, R_d)$ is a function of -mRNA concentration (m_i^-) and RdRp concentration (R_d), while for the -mRNA, this function $f(m_i^+, R_d)$ depends on +mRNA concentration (m_i^+) and RdRp concentration. This is because -mRNA acts as a template for the production of +mRNA using RdRp and vice versa (See Figure 4.2).

In the case of the -mRNA production there is no term representing normal transcription from a gene. This is because there is no gene that codes the -mRNA here. The only source for -mRNA is from RdRp transcription with +mRNA as a

template.

We considered the transcription of the mRNA using RdRp to be analogous to that of DNA-Dependent RNA Polymerase. The RdRp forms a complex with the mRNA. However, unlike in the transcription involving RNAPII, there is no open or closed complex here as the template here is an mRNA which is single-stranded. This process also involves abortive initiations where the RdRp starts transcription but aborts after a few nucleotides producing subgenomic RNA (sgmRNA). [146] These can be represented by the following set of equations:

$$m_i^+ + R_d \Longleftrightarrow (m_i^+ : R_d) \qquad m_i^- + R_d \Longleftrightarrow (m_i^- : R_d)$$

$$(m_i^+ : R_d) \longrightarrow m_i^+ + R_d + sgmRNA^- \qquad (m_i^- : R_d) \longrightarrow m_i^- + R_d + sgmRNA^+$$

$$(m_i^+ : R_d) \longrightarrow m_i^+ + R_d + m_i^- \qquad (m_i^- : R_d) \longrightarrow m_i^- + R_d + m_i^+$$

Consider the second set of reactions where +mRNA (m_i^+) is produced from -mRNA (m_i^-) and RdRp (R_d). The rate constants for these reactions can be denoted by :

$k_{E,i}^*$: Rate constant for transcription elongation

k_A : Rate constant for abortive initiation

k_+ & k_- : ON/OFF rate constants for RdRp-mRNA complex.

We assume that these are going to be the same for the RdRp interactions with both +mRNA and -mRNA. In this case, the rate of change of the complex concentration $(m_i^- : R_d)$ is given by the following equation :

$$\frac{\mathrm{d}(m_i^- : R_d)}{\mathrm{d}t} = k_+(m_i^-)(R_d) - k_-(m_i^- : R_d) - k_A(m_i^- : R_d) - k_{E,i}^*(m_i^- : R_d) \tag{4.7}$$

$$\text{At Steady State,} \quad (m_i^- : R_d) = K^{-1}(m_i^-)(R_d) \qquad \text{where } K^{-1} = \frac{k_+}{k_- + k_A + k_{E,i}^*}$$

In order to make this model more computationally simple, we are going to assume some amount of compartmentalization in the model. We assume that at pseudo-steady state conditions, RdRp is going to be equally divided between binding to -mRNA and +mRNA. This is will indeed be valid if the rate constants are the same for the interactions of RdRp with both RNAs. Assume that the total RdRp concentration available for binding with either one of these at a given time point = R_d^T, i.e., this is half the total amount of RdRp in the system.

$$R_d^T = R_d + (m_i^- : R_d) = R_d + K^{-1}(m_i^-)(R_d) \quad \Rightarrow \quad R_d = \frac{R_d^T}{1 + K^{-1}m_i^-} \tag{4.8}$$

$$f(m_i^-, R_d) = r_{X,i}^+ = k_{E,i}^*(m_i^- : R_d) = k_{E,i}^* m_i^- \left(\frac{R_d^T}{K + m_i^-}\right)$$

So, this new term can be replaced in equation (4.5) to give :

$$\dot{m}_i^+ = r_{X,i}\bar{u}_i - \theta_{m,i}m_i^+ + k_{E,i}^* m_i^- \left(\frac{R_d^T}{K + m_i^-}\right) \tag{4.9}$$

Similarly for -mRNA,

$$\dot{m}_i^- = -\theta_{m,i}m_i^- + k_{E,i}^* m_i^+ \left(\frac{R_d^T}{K + m_i^+}\right) \tag{4.10}$$

By solving the ODEs (4.9) and (4.10), it is possible to obtain the concentration of +mRNA 'i' at any point of time.

At Steady State,

$$m_i^{+*} = \frac{1}{\theta_{m,i}}\left[r_{X,i}\bar{u}_i + k_{E,i}^* m_i^{-*}\left(\frac{R_d^T}{K + m_i^{-*}}\right)\right] \qquad m_i^{-*} = \frac{1}{\theta_{m,i}}\left[k_{E,i}^* m_i^{+*}\left(\frac{R_d^T}{K + m_i^{+*}}\right)\right]$$

Here m_i^{+*} and m_i^{-*} represent the steady state concentrations of +mRNA and -mRNA respectively.

Combining these two equations, we get the following equation to denote the steady state concentration of the +mRNA 'i',

$$m_i^{+*} = \frac{1}{\theta_{m,i}} \left[r_{X,i}\bar{u}_i + k_{E,i}^* \left(\frac{1}{\theta_{m,i}} \left[k_{E,i}^* m_i^{+*} \left(\frac{R_d^T}{K + m_i^{+*}} \right) \right] \right) \left(\frac{R_d^T}{1 + \left(\frac{1}{\theta_{m,i}} \left[k_{E,i}^* m_i^{+*} \left(\frac{R_d^T}{K+m_i^{+*}} \right) \right] \right)} \right) \right]$$

We can use equations (4.9) and (4.10) to modify the concentration of +mRNA at any point of time. Then using this modified +mRNA concentration, we can use the normal set of translation equations with the same translation parameters to get the temporal variations in the concentration of the proteins. Since only the +mRNA concentration has been affected by the presence RdRp (the -mRNA doesn't code for any protein), that is the only factor that will affect the translation.

4.2.2 Control functions, Transcription and Translation Kinetic Limits

The control functions, w_i and u_i are the control functions modeled as:

$$\dot{w}_i = -\left(\frac{ln(2)}{\tau_{L,1/2}} \right) w_i \tag{4.11}$$

$$u_i = \frac{\sum_{j \in \chi} W_j f_j(...)}{\sum_{i \in C_i} W_i f_i(...)} \tag{4.12}$$

W_i denotes the weight of configuration i and f_i is the binding function modeled as a hill function. $\sum_{j \in \chi} W_j f_j(...)$ is the sum of all possible promoter configurations leading to transcription, while $\sum_{i \in C_i} W_i f_i(...)$ is the sum of all possible configurations for the Gene 'i'. The weights W_i are related to the Gibbs energy of

configuration 'i':

$$W_i = exp\left(\frac{-\Delta G_i}{RT}\right) \tag{4.13}$$

The rate of transcription $r_{X,j}u_j$ was modeled as the product of a kinetic limit, $r_{X,j}$, and a control term $u_j \in [0, 1]$. The kinetic limit of transcription was derived from elementary reactions leading to the formation of m_j, as previously done by McClure [101]. The kinetic limit of transcription was formulated as:

$$r_{X,j} = V_{X,j}^{max}\left(\frac{\mathcal{G}j}{\tau X, jK_{X,j} + \left(\tau_{X,j} + 1\right)\mathcal{G}_j}\right) \tag{4.14}$$

where the initiation time constant $\tau_{X,j}$, and saturation constant $K_{X,j}$ were experimentally measured by McClure [101]. The maximum transcription rate, $V_{X,j}^{max}$, was given by:

$$V_{X,j}^{max} = \left[R_{X,T}\left(\frac{\dot{v}X}{lG, j}\right)\right] \tag{4.15}$$

where $R_{X,T}$ is the RNA polymerase concentration (nM), $\dot{v}X$ is the RNA polymerase elongation rate (nt/h) and lG, j is the length of gene j in nucleotides (nt). By analogy, the kinetic limit of translation was formulated as:

$$r_{L,j} = V_{L,j}^{max}\left(\frac{x_{mRNA}}{\tau_{L,j}K_{L,j} + \left(\tau_{L,j} + 1\right)x_{mRNA}}\right) \tag{4.16}$$

where x_{mRNA} is the mRNA concentration and $V_{L,j}^{max}$ is the maximum translation rate which was formulated as:

$$V_{L,j}^{max} = \left[K_P R_{L,T}\left(\frac{\dot{v}L}{lP, j}\right)\right] \tag{4.17}$$

4.2.3 Building the model

A preliminary sequence-specific model was generated on Julia using the Gene Regulatory Network Generator - JuGRN [176]. This model takes into account

the transcription and translation parameters and solves for the protein expression as an FBA problem [181]. This model was modified to account for the effect of RdRp. The parameters for the model were obtained from BioNumbers or Vilkhovoy, et al [104,181]. Arbor Biosciences's myTXTL extract was chosen to be the cell-free extract used in this model. These model equations were solved numerically using the Rosenbrock23 routine of the DifferentialEquations.jl package [126]. The model code and parameters used in this chapter are available in the following Github repository: `https://github.coecis.cornell.edu/am3246/RdRp`

4.2.4 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 to synthesize nsp12 or nsp13 was 17 μL, composed of the Sigma 70 Master Mix (9 μL), T7 RNA polymerase DNA (1.5 μL - 0.225 nM), and the plasmids (6.5 μL - 7.65 nM) which were constructed with a T7 promoter. The working volume of all the reactions to synthesize nsp7 or nsp8 was 12.8 μL, composed of the Sigma 70 Master Mix (9 μL), GamS nuclease inhibitor (0.8 μL) and the linear DNA fragments (3 μL). These full constructs were ordered as linear DNA fragments, with 150-200 bp flanker sequences on both ends, from Twist Bioscience and Integrated DNA Technologies. The CFPS reactions were incubated at 29°C. The CFPS runs were carried out in triplicate for 2hr, 4hr, 8hr, and 16hr time points. Following each CFPS run, the samples were stored at -80°C.

4.2.5 mRNA Quantification

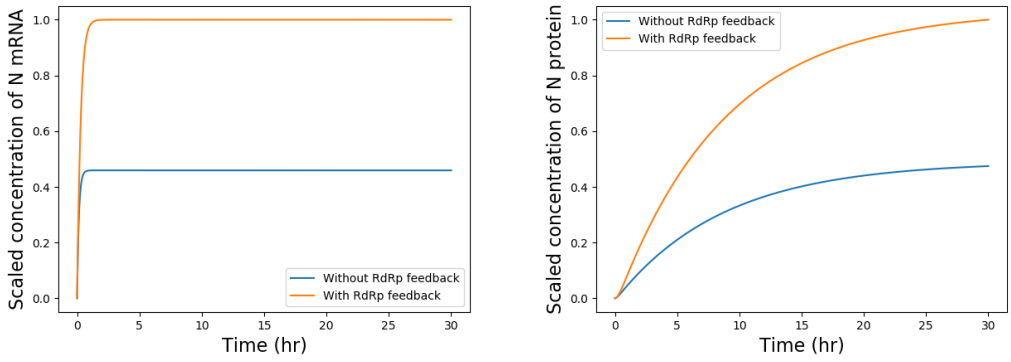
Following each CFPS run, the total RNA was extracted from 5 μL reaction mixture using PureLink RNA Mini Kit (Thermo Fisher Scientific) and stored at −80°C. The quantitative RT-PCR reactions were done using Applied Biosystems TaqMan RNA-to-CT 1-Step Kit and Custom TaqMan Gene Expression Assays (Thermo 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 DNA (nsp7, nsp8) were carried out for 2 hr. 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.

4.2.6 Protein Quantification

Following each CFPS run, 2 μL of each sample was extracted and added to 4 μL of water. And an equal amount of 2x Laemmli sample buffer (with 2-mercaptoethanol as a reducing agent) was added to these. An SDS-PAGE of 10μL of each of these diluted samples along with a Protein ladder of samples with known concentrations was performed at 300V until the dye fronts reached the bottom of the gel. This was followed by staining the gel with Bio-Safe™ Coomassie Stain (Bio-Rad Laboratories). The SDS-PAGE was performed

in a Mini-PROTEAN Tetra Cell (Bio-Rad Laboratories) using Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, 15 well, 15 µl gels (Bio-Rad Laboratories). The stained gel was visualized using the ChemiDoc MP and then the concentrations were measured using Image Lab software.

4.3 Results and Discussion



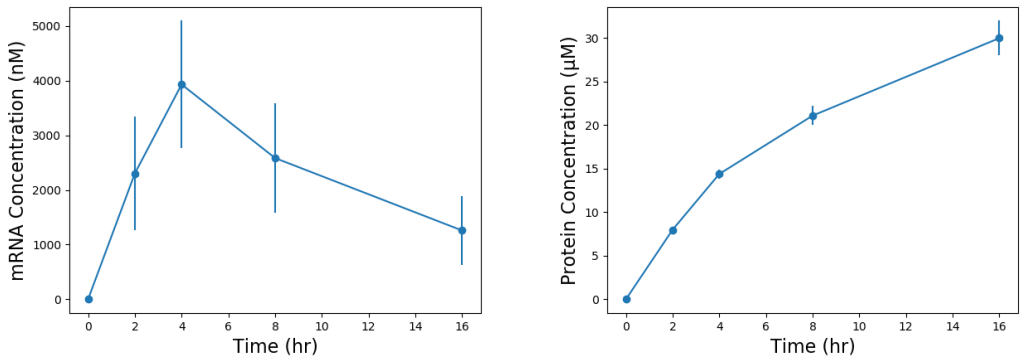
(a) (b)

Figure 4.3: Temporal variation of N mRNA and protein when expressed alongside RdRp. The blue curves represent the expression when the model doesn't account for the effect of RdRp, while the orange curves account for the RdRp feedback to amplify mRNA and thereby amplify protein expression

We used the developed model to simulate the effect of RdRp on an arbitrarily chosen SARS-CoV-2 protein with a comparatively low half-life - the nucleocapsid (N) protein of the SARS-CoV-2. The simulation used 5 nM each of DNA coding for nsp1 and RdRp. The simulation was run for a 30hr reaction time with two different models - One assumed a normal reaction where the effect of RdRp wasn't accounted for, and the other accounted for the RdRp's ability to amplify mRNA and thereby amplify protein expression. From Figure 4.3, we can observe that including the effect of RdRp on mRNA amplification does sig-

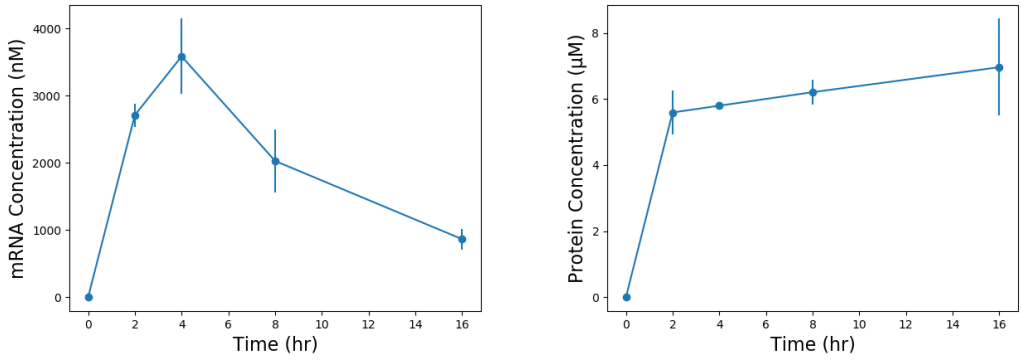
nificantly boost mRNA and thereby protein production. When we need to fit the model with experimental measurements, this model can easily be verified by replacing the chosen protein with a reporter protein such as deGFP.

We were able to synthesize the nsp7 (~ 10.2 kDa) and nsp8 (~ 22.8 kDa) proteins using the myTXTL cell-free system. With a starting concentration of 9.38nM of the nsp7/nsp8 DNA, we were able to obtain up to an average of 29.98 μM of nsp7 protein and 6.96 μM of nsp8 protein after a CFPS run of 16hr. The mRNA concentration peaked at the 4hr measurement point and continues to decline after that for both (Figures 4.4 and 4.5).



(a) (b)

Figure 4.4: Gene expression of nsp7. The points denote the mean, while the error bars denote one standard deviation calculated from three replicates



(a) (b)

Figure 4.5: Gene expression of nsp8. The points denote the mean, while the error bars denote one standard deviation calculated from three replicates

However, we were not able to synthesize the larger nsp12 (~ 106.2 kDa) and nsp13 (~ 67.8 kDa) subunits. Quantitative RT-PCR mRNA measurements of nsp12 and nsp13 CFPS samples after a 12 hour run revealed that only 0.58 nM of nsp12 mRNA and 2.97 nM of nsp13 mRNA was present. This explains the absence of protein and helps us narrow down the cause. We suspect the issue to be due to the presence of RNases (Ribonucleases) that are causing a higher degradation rate for mRNA than expected.

4.3.1 Future Direction

The model is still in its nascent stages and requires a lot of tweaking in order to make it more accurate. For one, currently, the model only takes into consideration the effect of RdRp and doesn't account for the formation of the RdRp complex along with nsp7, nsp8, and nsp13. The model will need to account for the parameters involved in the formation of this complex and the stability of the complex in addition to the stability of the individual subunits. We could

also explore building a model using FBA (Flux Balance Analysis) instead of the ODE-based model.

Secondly, several model parameters need to be determined using experimental data. The viral non-structural proteins involved with RdRp have already been successfully synthesized in *E. coli* cells [22,27,194]; So, it follows that it should still be theoretically possible to synthesize them in a *E. coli*-based cell-free extract. However, since we were unable to synthesize the larger nsp12 and nsp13 in our myTXTL system, we need to inspect other methods to synthesize them. Towards this aim, we can explore the use of other systems using extracts from other organisms such as the mammalian HeLa cell-free extract. The successful production of certain therapeutic proteins is often based on a synthesis setup that is closely related to in vivo conditions [199]. In the case it proves to be futile to synthesize the proteins in vivo, we could make use of traditional in vivo systems for the production of these proteins, but later assemble them for study in a cell-free system.

Lastly, once the binding parameters for the system are determined, it would be feasible to apply this model to cell-based systems to predict the spread of viral infection in human cells. If this approach is successful, it will shed light on the viral replication kinetics of all ssRNA viruses.

CHAPTER 5

CONCLUSION

We have constructed a promising transcriptional unit based on a gluconate-responsive allosteric transcription factor, GntR, that can be utilized to ”sense” glucose via its proxy, D-gluconate. The *E. coli* GntR combined with the mP70 promoter construct showed promising repression and de-repression results in the PURE system, 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. We aim to integrate this unit in a more complex genetic circuit to fully utilize its glucose-sensing potential to direct downstream synthesis of proteins such as insulin and glucagon for the management of diabetes. Second, we expressed the SARS-CoV-2 host shutoff factor, nsp1 in the myTXTL cell-free system and demonstrated that it does not bind the E.coli ribosomes. Thereby rendering its host translation inhibition ineffective in the *E.coli*-based cell-free system. Lastly, we also developed a mathematical model to study the effect of the SARS-CoV-2 RNA-dependent RNA polymerase (RdRp) on the production of a desired protein in the myTXTL cell-free system. Through simulations using the model built on a JuGRN backbone, we demonstrated that the presence of RdRp can amplify the mRNA in the system, thereby boosting protein production. We also postulate that RdRp can be harnessed to improve the production of desired proteins.

BIBLIOGRAPHY

[1] Diabetes facts and figures. https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html, 2019.

[2] 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, 2020.

[3] Syed Ovais Aftab, Muhammad Zubair Ghouri, Muhammad Umer Masood, Zeshan Haider, Zulqurnain Khan, Aftab Ahmad, and Nayla Munawar. Analysis of sars-cov-2 rna-dependent rna polymerase as a potential therapeutic drug target using a computational approach. *Journal of translational medicine*, 18(1):1–15, 2020.

[4] Dmitry E Agafonov, Vyacheslav A Kolb, Igor V Nazimov, and Alexander S Spirin. A protein residing at the subunit interface of the bacterial ribosome. *Proceedings of the National Academy of Sciences*, 96(22):12345–12349, 1999.

[5] Dae-Gyun Ahn, Jin-Kyu Choi, Deborah R Taylor, and Jong-Won Oh. Biochemical characterization of a recombinant sars coronavirus nsp12 rna-dependent rna polymerase capable of copying viral rna templates. *Archives of virology*, 157(11):2095–2104, 2012.

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

[8] 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.

[9] Maria Anastasina, Ilya Terenin, Sarah J Butcher, and Denis E Kainov. A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *Biotechniques*, 56(1):36–39, 2014.

[10] JC Atlas, EV Nikolaev, ST Browning, and ML Shuler. Incorporating genome-wide dna sequence information into a dynamic whole-cell model of escherichia coli: application to dna replication. *IET systems biology*, 2(5):369–382, 2008.

[11] Abhik K Banerjee, Mario R Blanco, Emily A Bruce, Drew D Honson, Linlin M Chen, Amy Chow, Prashant Bhat, Noah Ollikainen, Sofia A Quinodoz, Colin Loney, et al. Sars-cov-2 disrupts splicing, translation, and protein trafficking to suppress host defenses. *Cell*, 183(5):1325–1339, 2020.

[12] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Current opinion in genetics & development*, 15(2):116–124, 2005.

[13] Mahamaya Biswal, Stephen Diggs, Duo Xu, Nelli Khudaverdyan, Jiuwei Lu, Jian Fang, Gregor Blaha, Rong Hai, and Jikui Song. Two conserved oligomer interfaces of nsp7 and nsp8 underpin the dynamic assembly of sars-cov-2 rdrp. *Nucleic acids research*, 49(10):5956–5966, 2021.

[14] Igor W. Bogorad, Tzu-Shyang Lin, and James C. Liao. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature*, 502(7473):693–697, 2013.

[15] 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.

[16] Shelton S Bradrick, Simardeep Nagyal, and Hilary Novatt. A mirna-responsive cell-free translation system facilitates isolation of hepatitis c virus mirnp complexes. *Rna*, 19(8):1159–1169, 2013.

[17] 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.

[18] Kara A. Calhoun and James R. Swartz. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnology and Bioengineering*, 90(5):606–613, 2005.

[19] Mariajose Castellanos, David B Wilson, and Michael L Shuler. A modular minimal cell model: purine and pyrimidine transport and metabolism. *Proceedings of the National Academy of Sciences*, 101(17):6681–6686, 2004.

[20] 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.

[21] James Chappell, Melissa K Takahashi, and Julius B Lucks. Creating small transcription activating rnas. *Nature chemical biology*, 11(3):214–220, 2015.

[22] James Chen, Brandon Malone, Eliza Llewellyn, Michael Grasso, Patrick MM Shelton, Paul Dominic B Olinares, Kashyap Maruthi, Edward T Eng, Hasan Vatandaslar, Brian T Chait, et al. Structural basis for helicase-polymerase coupling in the sars-cov-2 replication-transcription complex. *Cell*, 182(6):1560–1573, 2020.

[23] 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.

[24] David Dai, Nicholas Horvath, and Jeffrey Varner. Dynamic sequence specific constraint-based modeling of cell-free protein synthesis. *Processes*, 6(8):132, 2018.

[25] Samar Damiati, Rami Mhanna, Rimantas Kodzius, and Eva-Kathrin Ehmoser. Cell-free approaches in synthetic biology utilizing microfluidics. *Genes*, 9(3):144, 2018.

[26] Denis Daneman. Type 1 diabetes. *The Lancet*, 367(9513):847–858, 2006.

[27] Tyler L Dangerfield, Nathan Z Huang, and Kenneth A Johnson. Expression and purification of tag-free sars-cov-2 rna-dependent rna polymerase in escherichia coli. *STAR protocols*, 2(1):100357, 2021.

[28] 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*, 2018:7021826–7021826, 2018.

[29] Mark R Denison, Rachel L Graham, Eric F Donaldson, Lance D Eckerle, and Ralph S Baric. Coronaviruses: an rna proofreading machine regulates replication fidelity and diversity. *RNA biology*, 8(2):270–279, 2011.

[30] Devasenan Devendra, Edwin Liu, and George S Eisenbarth. Type 1 diabetes: recent developments. *Bmj*, 328(7442):750–754, 2004.

[31] MM Domach, SK Leung, RE Cahn, GG Cocks, and ML Shuler. Computer model for glucose-limited growth of a single cell of escherichia coli b/r-a. *Biotechnology and bioengineering*, 26(9):1140–1140, 1984.

[32] 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.

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

[34] 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):1–10, 2000.

[35] Toru Ezure, Takashi Suzuki, Shoken Higashide, Eiichi Shintani, Kohki Endo, Shin-ichiro Kobayashi, Masamitsu Shikata, Masaaki Ito, Koji Tanimizu, and Osamu Nishimura. Cell-free protein synthesis system prepared from insect cells by freeze-thawing. *Biotechnology progress*, 22(6):1570–1577, 2006.

[36] Anthony R Fehr and Stanley Perlman. Coronaviruses: an overview of their replication and pathogenesis. *Coronaviruses*, pages 1–23, 2015.

[37] Yaara Finkel, Orel Mizrahi, Aharon Nachshon, Shira Weingarten-Gabbay, David Morgenstern, Yfat Yahalom-Ronen, Hadas Tamir, Hagit Achdout, Dana Stein, Ofir Israeli, et al. The coding capacity of sars-cov-2. *Nature*, 589(7840):125–130, 2021.

[38] AG Fredrickson. Formulation of structured growth models. *Biotechnology and bioengineering*, 18(10):1481–1486, 1976.

[39] Y Fujita and Y Miwa. Identification of an operator sequence for the bacillus subtilis gnt operon. *Journal of Biological Chemistry*, 264(7):4201–4206, 1989.

[40] Ernest F Gale, Joan P Folkes, et al. Effect of nucleic acids on protein syn-

thesis and amino-acid incorporation in disrupted staphylococcal cells. *Nature*, 173:1223–7, 1954.

[41] Yan Gao, Liming Yan, Yucen Huang, Fengjiang Liu, Yao Zhao, Lin Cao, Tao Wang, Qianqian Sun, Zhenhua Ming, Lianqi Zhang, et al. Structure of the rna-dependent rna polymerase from covid-19 virus. *Science*, 368(6492):779–782, 2020.

[42] 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.

[43] V. Georgi, Leopold Georgi, Martin Blechert, Merlin Bergmeister, Michael Zwanzig, Doreen A. Wüstenhagen, Frank F. Bier, Erik Jung, and Stefan 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, 2016.

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

[45] Calvin J Gordon, Egor P Tchesnokov, Joy Y Feng, Danielle P Porter, and Matthias Götte. The antiviral compound remdesivir potently inhibits rna-dependent rna polymerase from middle east respiratory syndrome coronavirus. *Journal of Biological Chemistry*, 295(15):4773–4779, 2020.

[46] 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.

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

[48] 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.

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

[50] Andras Gyorgy and Richard 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. IEEE, 2016.

[51] 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.

[52] Joshua J Hamilton, Vivek Dwivedi, and Jennifer L Reed. Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophysical journal*, 105(2):512–522, 2013.

[53] 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*, 50(33):7502–7519, 2011.

[54] Hartmut Hengel, Ulrich H Koszinowski, and Karl-Klaus Conzelmann. Viruses know it all: new insights into ifn networks. *Trends in immunology*, 26(7):396–401, 2005.

[55] Christopher S Henry, Linda J Broadbelt, and Vassily Hatzimanikatis. Thermodynamics-based metabolic flux analysis. *Biophysical journal*, 92(5):1792–1805, 2007.

[56] 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):1–9, 2008.

[57] Hauke S Hillen, Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov, and Patrick Cramer. Structure of replicating sars-cov-2 polymerase. *Nature*, 584(7819):154–156, 2020.

[58] Yu HongFeng, Ding YongQian, Liu HaiTao, Zhu WenQian, Liu Guo-Qiang, Fu XiuQing, and Ding WeiMin. Optimization design and application of variable rate fertilization system for small-scaled fields. *Transactions of the Chinese Society of Agricultural Engineering*, 34(3):35–41, 2018.

[59] 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. *Metabolic engineering communications*, 10:e00113, 2020.

[60] Chelsea Y Hu, Jeffrey D Varner, and Julius B Lucks. Generating effective models and parameters for rna genetic circuits. *ACS synthetic biology*, 4(8):914–926, 2015.

[61] Cheng Huang, Kumari G Lokugamage, Janet M Rozovics, Krishna Narayanan, Bert L Semler, and Shinji Makino. Sars coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mrnas: viral mrnas are resistant to nsp1-induced rna cleavage. *PLoS pathogens*, 7(12):e1002433, 2011.

[62] Isabelle Imbert, Jean-Claude Guillemot, Jean-Marie Bourhis, Cécile Bussetta, Bruno Coutard, Marie-Pierre Egloff, François Ferron, Alexander E Gorbalenya, and Bruno Canard. A second, non-canonical rna-dependent rna polymerase in sars coronavirus. *The EMBO journal*, 25(20):4933–4942, 2006.

[63] Konstantin A Ivanov, Volker Thiel, Jessika C Dobbe, Yvonne Van Der Meer, Eric J Snijder, and John Ziebuhr. Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *Journal of virology*, 78(11):5619–5632, 2004.

[64] Konstantin A Ivanov and John Ziebuhr. Human coronavirus 229e nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and rna 5-triphosphatase activities. *Journal of virology*, 78(14):7833–7838, 2004.

[65] 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):1–11, 2018.

[66] Andrew R Jauregui, Dhruti Savalia, Virginia K Lowry, Cara M Farrell, and Marc G Wathelet. Identification of residues of sars-cov nsp1 that differentially affect inhibition of gene expression and antiviral signaling. *PloS one*, 8(4):e62416, 2013.

[67] 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. *Molecular systems biology*, 4(1):220, 2008.

[68] 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.

[69] 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.

[70] Guoqiang Jiang and Bei B Zhang. Glucagon and regulation of glucose metabolism. *American Journal of Physiology-Endocrinology And Metabolism*, 284(4):E671–E678, 2003.

[71] Yi Jiang, Wanchao Yin, and H Eric Xu. Rna-dependent rna polymerase: Structure, mechanism, and drug discovery for covid-19. *Biochemical and biophysical research communications*, 538:47–53, 2021.

[72] Jaeyoung K Jung, Khalid K Alam, Matthew S Verosloff, Daiana A Capdevila, Morgane Desmau, Phillip R Clauer, Jeong Wook Lee, Peter Q Nguyen, Pablo A Pastén, Sandrine J Matiasek, et al. Cell-free biosensors for rapid detection of water contaminants. *Nature biotechnology*, 38(12):1451–1459, 2020.

[73] Wataru Kamitani, Cheng Huang, Krishna Narayanan, Kumari G Lokugamage, and Shinji Makino. A two-pronged strategy to suppress host protein synthesis by sars coronavirus nsp1 protein. *Nature structural & molecular biology*, 16(11):1134–1140, 2009.

[74] 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, 2011.

[75] Rozhgar A Khailany, Muhamad Safdar, and Mehmet Ozaslan. Genomic characterization of a novel sars-cov-2. *Gene reports*, 19:100682, 2020.

[76] Akram T Kharroubi and Hisham M Darwish. Diabetes mellitus: The epidemic of the century. *World Journal of Diabetes*, 6(6):850–867, 2015.

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

[78] Dongwan Kim, Joo-Yeon Lee, Jeong-Sun Yang, Jun Won Kim, V Narry

Kim, and Hyeshik Chang. The architecture of sars-cov-2 transcriptome. *Cell*, 181(4):914–921, 2020.

[79] Robert N Kirchdoerfer and Andrew B Ward. Structure of the sars-cov nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications*, 10(1):1–9, 2019.

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

[81] Goran Kokic, Hauke S Hillen, Dimitry Tegunov, Christian Dienemann, Florian Seitz, Jana Schmitzova, Lucas Farnung, Aaron Siewert, Claudia Höbartner, and Patrick Cramer. Mechanism of sars-cov-2 polymerase stalling by remdesivir. *Nature communications*, 12(1):1–7, 2021.

[82] 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.

[83] Christopher P Lapointe, Rosslyn Grosely, Alex G Johnson, Jinfan Wang, Israel S Fernández, and Joseph D Puglisi. Dynamic competition between sars-cov-2 nsp1 and mrna on the human ribosome inhibits translation initiation. *Proceedings of the National Academy of Sciences*, 118(6), 2021.

[84] 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.

[85] Na-Ra Lee, Hyun-Mi Kwon, Kkothanahreum Park, Sangtaek Oh, Yong-Joo Jeong, and Dong-Eun Kim. Cooperative translocation enhances the unwinding of duplex dna by sars coronavirus helicase nsp13. *Nucleic acids research*, 38(21):7626–7636, 2010.

[86] Kathleen C Lehmann, Anastasia Gulyaeva, Jessika C Zevenhoven-Dobbe, George MC Janssen, Mark Ruben, Hermen S Overkleeft, Peter A Van Veelen, Dmitry V Samborskiy, Alexander A Kravchenko, Andrey M Leontovich, et al. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the rna polymerase-containing protein of all nidoviruses. *Nucleic acids research*, 43(17):8416–8434, 2015.

[87] 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*, 2(9):482–489, 2013.

[88] Nathan E Lewis, Harish Nagarajan, and Bernhard O Palsson. Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods. *Nature Reviews Microbiology*, 10(4):291–305, 2012.

[89] Dongsheng Li, Ting Wei, Catherine M Abbott, and David Harrich. The unexpected roles of eukaryotic translation elongation factors in rna virus replication and pathogenesis. *Microbiol. Mol. Biol. Rev.*, 77(2):253–266, 2013.

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

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

[92] Yuan Lu. Cell-free synthetic biology: Engineering in an open world. *Synthetic and systems biotechnology*, 2(1):23–27, 2017.

[93] 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. *Proceedings of the National Academy of Sciences*, 111(1):125–130, 2014.

[94] Julius B Lucks, Lei Qi, Vivek K Mutalik, Denise Wang, and Adam P Arkin. Versatile rna-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*, 108(21):8617–8622, 2011.

[95] Guangxiang Luo, Robert K Hamatake, Danielle M Mathis, Jason Racela, Karen L Rigat, Julie Lemm, and Richard J Colonno. De novo initiation of rna synthesis by the rna-dependent rna polymerase (ns5b) of hepatitis c virus. *Journal of virology*, 74(2):851–863, 2000.

[96] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in escherichia coli via the lacr/o, the tetr/o and arac/i1-i2 regulatory elements. *Nucleic acids research*, 25(6):1203–1210, 1997.

[97] 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), 2018.

[98] Brandon Malone, James Chen, Qi Wang, Eliza Llewellyn, Young Joo Choi, Paul Dominic B Olinares, Xinyun Cao, Carolina Hernandez, Edward T Eng, Brian T Chait, et al. Structural basis for backtracking by the sars-cov-2 replication–transcription complex. *Proceedings of the National Academy of Sciences*, 118(19), 2021.

[99] 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, Ashleigh E. Stoddart, Miriam Amiram, Arnaz K. Ranji Charna, Jaymin R. Patel, Farren J. Isaacs, Neil L. Kelleher, Seok Hoon Hong, and Michael C. Jewett. Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nature Communications*, 9(1):1203, 2018.

[100] J Heinrich Matthaei and Marshall W Nirenberg. Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. *Proceedings of the National Academy of Sciences of the United States of America*, 47(10):1580, 1961.

[101] William R McClure. Rate-limiting steps in rna chain initiation. *Proceedings of the National Academy of Sciences*, 77(10):5634–5638, 1980.

[102] D. T. McQuade and Peter H. Seeberger. Applying flow chemistry: Methods, materials, and multistep synthesis. *Journal of Organic Chemistry*, 78(13):6384–6389, 2013.

[103] Satoshi Mikami, Tominari Kobayashi, and Hiroaki Imataka. Cell-free protein synthesis systems with extracts from cultured human cells. In *Cell-Free Protein Production*, pages 43–52. Springer, 2010.

[104] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1):D750–D753, 2010.

[105] Ekaterina Minskaia, Tobias Hertzig, Alexander E Gorbalenya, Valérie Campanacci, Christian Cambillau, Bruno Canard, and John Ziebuhr. Discovery of an rna virus 3→ 5 exoribonuclease that is critically involved in coronavirus rna synthesis. *Proceedings of the National Academy of Sciences*, 103(13):5108–5113, 2006.

[106] Simon J. Moore, James T. MacDonald, Sarah Wienecke, Alka Ishwarbhai, Argyro Tsipa, Rochelle Aw, Nicolas Kylilis, David J. Bell, David W. McClymont, Kirsten Jensen, Karen M. Polizzi, Rebekka Biedendieck, and Paul S. Freemont. Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19):201715806, 2018.

[107] 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):13, 2019.

[108] Keisuke Nakagawa, Krishna Narayanan, Masami Wada, Vsevolod L Popov, Maria Cajimat, Ralph S Baric, and Shinji Makino. The endonucleolytic rna cleavage function of nsp1 of middle east respiratory syndrome coronavirus promotes the production of infectious virus particles in specific human cell lines. *Journal of virology*, 92(21):e01157–18, 2018.

[109] Ahmad Abu Turab Naqvi, Kisa Fatima, Taj Mohammad, Urooj Fatima, Indrakant K Singh, Archana Singh, Shaikh Muhammad Atif, Gururao Hariprasad, Gulam Mustafa Hasan, and Md Imtaiyaz Hassan. Insights into sars-cov-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1866(10):165878, 2020.

[110] Krishna Narayanan, Cheng Huang, Kumari Lokugamage, Wataru Kamitani, Tetsuro Ikegami, Chien-Te K Tseng, and Shinji Makino. Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type i interferon, in infected cells. *Journal of virology*, 82(9):4471–4479, 2008.

[111] Krishna Narayanan, Sydney I Ramirez, Kumari G Lokugamage, and Shinji Makino. Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus research*, 202:89–100, 2015.

[112] Katerina Naydenova, Kyle W Muir, Long-Fei Wu, Ziguo Zhang, Francesca Coscia, Mathew J Peet, Pablo Castro-Hartmann, Pu Qian, Kasim Sader, Kyle Dent, et al. Structure of the sars-cov-2 rna-dependent rna polymerase in the presence of favipiravir-rtp. *Proceedings of the National Academy of Sciences*, 118(7), 2021.

[113] 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. *Proceedings of the National Academy of Sciences*, 109(36):14526–14531, 2012.

[114] Pamela N. Nge, Chad I. Rogers, and Adam T. Woolley. Advances in microfluidic materials, functions, integration, and applications. *Chemical Reviews*, 113(4):2550–2583, 2013.

[115] 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, 2013.

[116] 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, 2017.

[117] Marshall W Nirenberg and J Heinrich Matthaei. The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences*, 47(10):1588–1602, 1961.

[118] Vincent Noireaux, Roy Bar-Ziv, and Albert Libchaber. Principles of cell-free genetic circuit assembly. *Proceedings of the National Academy of Sciences*, 100(22):12672–12677, 2003.

[119] 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. *Molecular systems biology*, 9(1):693, 2013.

[120] Paul H Opgenorth, Tyler P Korman, and James U Bowie. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature communications*, 5(1):1–8, 2014.

[121] Keith Pardee and James J. Collins. Paper-based synthetic gene networks, 2014.

[122] Keith Pardee, Alexander A. Green, Melissa K. Takahashi, Dana Braff, Guillaume Lambert, Jeong Wook Lee, Tom Ferrante, Duo Ma, Nina Donghia, Melina Fan, Nichole M. Daringer, Irene Bosch, Dawn M. Dudley, David H. O'Connor, Lee Gehrke, and James J. Collins. Rapid, low-cost

detection of zika virus using programmable biomolecular components. *Cell*, 165(5):1255–1266, 2016.

[123] Susan Payne. Introduction to rna viruses. *Viruses*, page 97, 2017.

[124] Norbert Peekhaus and T. Conway. Positive and negative transcriptional regulation of the escherichia coli gluconate regulon gene gntt by gntr and the cyclic amp (camp)-camp receptor protein complex. *Journal of Bacteriology*, 180(7):1777–1785, 1998.

[125] Clara C Posthuma, Aartjan JW Te Velthuis, and Eric J Snijder. Nidovirus rna polymerases: complex enzymes handling exceptional rna genomes. *Virus research*, 234:58–73, 2017.

[126] Aditya Pratapa, Amogh P Jalihal, Jeffrey N Law, Aditya Bharadwaj, and TM Murali. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature methods*, 17(2):147–154, 2020.

[127] JM Pratt. Coupled transcription-translation in prokaryotic cell-free systems. *Transcription and translation: a practical approach*, 1984.

[128] Sephra Rampersad and Paula Tennant. Replication and expression strategies of viruses. *Viruses*, page 55, 2018.

[129] Jan Rehwinkel and Michaela U Gack. Rig-i-like receptors: their regulation and roles in rna sensing. *Nature Reviews Immunology*, 20(9):537–551, 2020.

[130] 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.

[131] Maria Romano, Alessia Ruggiero, Flavia Squeglia, Giovanni Maga, and Rita Berisio. A structural view of sars-cov-2 rna replication machinery: Rna synthesis, proofreading and final capping. *Cells*, 9(5):1267, 2020.

[132] 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.

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

[134] Shun Sakuraba, Xie Qilin, Kota Kasahara, Junichi Iwakiri, and Hidetoshi Kono. Modeling sars-cov-2 nsp1-5′-utr complex via the extended ensemble simulations. *bioRxiv*, 2021.

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

[136] Inga-Marie Schaefer, Robert F Padera, Isaac H Solomon, Sanjat Kanjilal, Mark M Hammer, Jason L Hornick, and Lynette M Sholl. In situ detection of sars-cov-2 in lungs and airways of patients with covid-19. *Modern Pathology*, 33(11):2104–2114, 2020.

[137] Christophe H Schilling, David Letscher, and Bernhard Ø Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of theoretical biology*, 203(3):229–248, 2000.

[138] Katharina Schubert, Evangelos D Karousis, Ahmad Jomaa, Alain Scaiola, Blanca Echeverria, Lukas-Adrian Gurzeler, Marc Leibundgut, Volker Thiel, Oliver Mühlemann, and Nenad Ban. Sars-cov-2 nsp1 binds the ribosomal mrna channel to inhibit translation. *Nature structural & molecular biology*, 27(10):959–966, 2020.

[139] Stefan Schuster, David A Fell, and Thomas Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature biotechnology*, 18(3):326–332, 2000.

[140] Anja Seybert, Annette Hegyi, Stuart G Siddell, and John Ziebuhr. The human coronavirus 229e superfamily 1 helicase has rna and dna duplex-unwinding activities with 5-to-3 polarity. *Rna*, 6(7):1056–1068, 2000.

[141] Jian Shang, Yushun Wan, Chuming Luo, Gang Ye, Qibin Geng, Ashley Auerbach, and Fang Li. Cell entry mechanisms of sars-cov-2. *Proceedings of the National Academy of Sciences*, 117(21):11727–11734, 2020.

[142] 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):751–755, 2001.

[143] 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):1–9, 2010.

[144] 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.

[145] Takehiro Shinoda, Naoko Shinya, Kaori Ito, Yoshiko Ishizuka-Katsura, Noboru Ohsawa, Takaho Terada, Kunio Hirata, Yoshiaki Kawano, Masaki Yamamoto, Taisuke Tomita, Yohei Ishibashi, Yoshio Hirabayashi, Tomomi Kimura-Someya, Mikako Shirouzu, and Shigeyuki Yokoyama. Cell-free methods to produce structurally intact mammalian membrane proteins. *Scientific Reports*, 6(1):30442, 2016.

[146] Bo Shu and Peng Gong. Structural basis of viral rna-dependent rna polymerase catalysis and translocation. *Proceedings of the National Academy of Sciences*, 113(28):E4005–E4014, 2016.

[147] Robert W Siegel, Scott Adkins, and C Cheng Kao. Sequence-specific recognition of a subgenomic rna promoter by a viral rna polymerase. *Proceedings of the National Academy of Sciences*, 94(21):11238–11243, 1997.

[148] Nicolas Sierro, Yuko Makita, Michiel J. L. de Hoon, and Kenta Nakai. Dbtbs: a database of transcriptional regulation in bacillus subtilis containing upstream intergenic conservation information. *Nucleic Acids Research*, 36:93–96, 2008.

[149] 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, 2020.

[150] Matthieu Simeoni, Théo Cavinato, Daniel Rodriguez, and David Gatfield. I (nsp1) ecting sars-cov-2–ribosome interactions. *Communications biology*, 4(1):1–5, 2021.

[151] Everett Clinton Smith, Nicole R Sexton, and Mark R Denison. Thinking outside the triangle: replication fidelity of the largest rna viruses. *Annual review of virology*, 1:111–132, 2014.

[152] EJ Snijder, E Decroly, and J Ziebuhr. The nonstructural proteins directing coronavirus rna synthesis and processing. *Advances in virus research*, 96:59–126, 2016.

[153] 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.

[154] Konstantin MJ Sparrer and Michaela U Gack. Intracellular detection of viral nucleic acids. *Current opinion in microbiology*, 26:1–9, 2015.

[155] Alexander S Spirin, Vladimir I Baranov, Lubov A Ryabova, Sergey Yu Ovodov, and Yuly B Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162–1164, 1988.

[156] DE Steinmeyer and ML Shuler. Structured model for saccharomyces cerevisiae. *Chemical engineering science*, 44(9):2017–2030, 1989.

[157] 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.

[158] Lorenzo Subissi, Clara C Posthuma, Axelle Collet, Jessika C Zevenhoven-Dobbe, Alexander E Gorbalenya, Etienne Decroly, Eric J Snijder, Bruno Canard, and Isabelle Imbert. One severe acute respiratory syndrome coronavirus protein complex integrates processive rna polymerase and exonuclease activities. *Proceedings of the National Academy of Sciences*, 111(37):E3900–E3909, 2014.

[159] Rahul K Suryawanshi, Raghuram Koganti, Alex Agelidis, Chandrashekhar D Patil, and Deepak Shukla. Dysregulation of cell signaling by sars-cov-2. *Trends in microbiology*, 2020.

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

[161] Stanley Tabor. Expression using the t7 rna polymerase/promoter system. *Current protocols in molecular biology*, 11(1):16–2, 1990.

[162] Melissa K Takahashi, James Chappell, Clarmyra A Hayes, Zachary Z Sun, Jongmin Kim, Vipul Singhal, Kevin J Spring, Shaima Al-Khabouri, Christopher P Fall, Vincent Noireaux, et al. Rapidly characterizing the fast dynamics of rna genetic circuitry with cell-free transcription–translation (tx-tl) systems. *ACS synthetic biology*, 4(5):503–515, 2015.

[163] Melissa K Takahashi, Clarmyra A Hayes, James Chappell, Zachary Z Sun, Richard M Murray, Vincent Noireaux, and Julius B Lucks. Characterizing and prototyping genetic networks with cell-free transcription–translation reactions. *Methods*, 86:60–72, 2015.

[164] Kazuyuki Takai, Tatsuya Sawasaki, and Yaeta Endo. Practical cell-free protein synthesis system using purified wheat embryos. *Nature protocols*, 5(2):227–238, 2010.

[165] Tomohisa Tanaka, Wataru Kamitani, Marta L DeDiego, Luis Enjuanes, and Yoshiharu Matsuura. Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mrna. *Journal of virology*, 86(20):11128–11137, 2012.

[166] Matthew Zirui Tay, Chek Meng Poh, Laurent Rénia, Paul A MacAry, and Lisa FP Ng. The trinity of covid-19: immunity, inflammation and intervention. *Nature Reviews Immunology*, 20(6):363–374, 2020.

[167] Noah D Taylor, Alexander S Garruss, Rocco Moretti, Sum Chan, Mark A Arbing, Duilio Cascio, Jameson K Rogers, Farren J Isaacs, Sriram Kosuri, David Baker, et al. Engineering an allosteric transcription factor to respond to new ligands. *Nature methods*, 13(2):177–183, 2016.

[168] Aartjan JW Te Velthuis, Sjoerd HE Van Den Worm, and Eric J Snijder. The sars-coronavirus nsp7+ nsp8 complex is a unique multimeric rna polymerase capable of both de novo initiation and primer extension. *Nucleic acids research*, 40(4):1737–1747, 2012.

[169] Matthias Thoms, Robert Buschauer, Michael Ameismeier, Lennart Koepke, Timo Denk, Maximilian Hirschenberger, Hanna Kratzat, Manuel Hayn, Timur Mackens-Kiani, Jingdong Cheng, et al. Structural basis for translational shutdown and immune evasion by the nsp1 protein of sars-cov-2. *Science*, 369(6508):1249–1255, 2020.

[170] Aidan Tinafar, Katariina Jaenes, and Keith Pardee. Synthetic biology goes cell-free. *BMC biology*, 17(1):1–14, 2019.

[171] Masaru Tomita, Kenta Hashimoto, Koichi Takahashi, Thomas Simon Shimizu, Yuri Matsuzaki, Fumihiko Miyoshi, Kanako Saito, Sakura Tanida, Katsuyuki Yugi, J Craig Venter, et al. E-cell: software environment for whole-cell simulation. *Bioinformatics (Oxford, England)*, 15(1):72–84, 1999.

[172] Chien-Te K Tseng, Jennifer Tseng, Lucy Perrone, Melissa Worthy, Vsevolod Popov, and Clarence J Peters. Apical entry and release of severe acute respiratory syndrome-associated coronavirus in polarized calu-3 lung epithelial cells. *Journal of virology*, 79(15):9470–9479, 2005.

[173] 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*, 2017:9529501–9529501, 2017.

[174] 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):565–565, 2019.

[175] Amit Varma and Bernhard O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Applied and environmental microbiology*, 60(10):3724–3731, 1994.

[176] Jeffrey D. Varner. Gene Regulatory Network Generator in Julia (JuGRN), April 2020.

[177] Jeffrey D Varner. Lectures notes in advanced biomolecular engineering, 2020.

[178] Sangita Venkataraman, Burra VLS Prasad, and Ramasamy Selvarajan. Rna dependent rna polymerases: insights from structure, function and evolution. *Viruses*, 10(2):76, 2018.

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

[180] 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. *JoVE (Journal of Visualized Experiments)*, (152):e60329, 2019.

[181] Michael Vilkhovoy, Nicholas Horvath, Che-Hsiao Shih, Joseph A Wayman, Kara Calhoun, James Swartz, and Jeffrey D Varner. Sequence spe-

cific modeling of e. coli cell-free protein synthesis. *ACS synthetic biology*, 7(8):1844–1857, 2018.

[182] Peter L. Voyvodic and Jerome Bonnet. Cell-free biosensors for biomedical applications. *Current Opinion in Biomedical Engineering*, 13:9–15, 2020.

[183] Philip V'kovski, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. Coronavirus biology and replication: implications for sars-cov-2. *Nature Reviews Microbiology*, 19(3):155–170, 2021.

[184] Huihui Wang, Xuemei Li, Tao Li, Shubing Zhang, Lianzi Wang, Xian Wu, and Jiaqing Liu. The genetic sequence, origin, and diagnosis of sars-cov-2. *European Journal of Clinical Microbiology & Infectious Diseases*, pages 1–7, 2020.

[185] 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. *Proceedings of the National Academy of Sciences*, 108(22):9049–9054, 2011.

[186] Joseph A Wayman, Adithya Sagar, and Jeffrey D Varner. Dynamic modeling of cell-free biochemical networks using effective kinetic models. *Processes*, 3(1):138–160, 2015.

[187] LA Weber, ER Feman, and C Baglioni. Cell free system from hela cells active in initiation of protein synthesis. *Biochemistry*, 14(24):5315–5321, 1975.

[188] Michael Weinfeld, Jane Lee, Gu Ruiqi, Feridoun Karimi-Busheri, Davis Chen, and Joan Allalunis-Turner. Use of a postlabelling assay to examine the removal of radiation-induced dna lesions by purified enzymes and human cell extracts. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 378(1-2):127–137, 1997.

[189] Wolfgang Wiechert. 13c metabolic flux analysis. *Metabolic engineering*, 3(3):195–206, 2001.

[190] T Winnick et al. Studies on the mechanism of protein, synthesis in embryonic and tumor tissues. 1. evidence relating to the incorporation of labeled amino acids into protein structure in homogenates. *Arch. Biochem.*, 27, 1950.

[191] Aiping Wu, Yousong Peng, Baoying Huang, Xiao Ding, Xianyue Wang, Peihua Niu, Jing Meng, Zhaozhong Zhu, Zheng Zhang, Jiangyuan Wang, et al. Genome composition and divergence of the novel coronavirus (2019-ncov) originating in china. *Cell host & microbe*, 27(3):325–328, 2020.

[192] P Wu, NG Ray, and ML Shuler. A single-cell model for cho cells a. *Annals of the New York Academy of Sciences*, 665(1):152–187, 1992.

[193] Raghavendra Yadavalli and Tobili Sam-Yellowe. Hela based cell free expression systems for expression of plasmodium rhoptry proteins. *JoVE (Journal of Visualized Experiments)*, (100):e52772, 2015.

[194] Liming Yan, Ying Zhang, Ji Ge, Litao Zheng, Yan Gao, Tao Wang, Zhihui Jia, Haofeng Wang, Yucen Huang, Mingyu Li, et al. Architecture of a sars-cov-2 mini replication and transcription complex. *Nature communications*, 11(1):1–6, 2020.

[195] Wanchao Yin, Chunyou Mao, Xiaodong Luan, Dan-Dan Shen, Qingya Shen, Haixia Su, Xiaoxi Wang, Fulai Zhou, Wenfeng Zhao, Minqi Gao, et al. Structural basis for inhibition of the rna-dependent rna polymerase from sars-cov-2 by remdesivir. *Science*, 368(6498):1499–1504, 2020.

[196] Ken-Ichi Yoshida, Yasutaro Fujita, and Akinori Sarai. Missense mutations in the bacillus subtilis gnt repressor that diminish operator binding ability, 1993.

[197] Ken-ichi Yoshida, Hisanobu Ohmori, Yasuhiko Miwa, and Yasutaro Fujita. Bacillus subtilis gnt repressor mutants that diminish gluconate-binding ability. *Journal of bacteriology*, 177(16):4813–4816, 1995.

[198] Francis K Yoshimoto. The proteins of severe acute respiratory syndrome coronavirus-2 (sars cov-2 or n-cov19), the cause of covid-19. *The protein journal*, 39:198–216, 2020.

[199] Anne Zemella, Lena Thoring, Christian Hoffmeister, and Stefan Kubick. Cell-free protein synthesis: pros and cons of prokaryotic and eukaryotic systems. *ChemBioChem*, 16(17):2420, 2015.

[200] Yang Zhang. Genome wide structure and function modeling of sars-cov-2. `https://zhanglab.ccmb.med.umich.edu/COVID-19/`, 2020.

[201] Ying Zhang, Ines Thiele, Dana Weekes, Zhanwen Li, Lukasz Jaroszewski,

Krzysztof Ginalski, Ashley M Deacon, John Wooley, Scott A Lesley, Ian A Wilson, et al. Three-dimensional structural view of the central metabolic network of thermotoga maritima. *Science*, 325(5947):1544–1549, 2009.

[202] Wei Zhu, Catherine Z Chen, Kirill Gorshkov, Miao Xu, Donald C Lo, and Wei Zheng. Rna-dependent rna polymerase as a target for covid-19 drug discovery. *SLAS DISCOVERY: Advancing the Science of Drug Discovery*, 25(10):1141–1151, 2020.