



Effective Biophysical Modeling of Cell Free Transcription and Translation Processes

Abhinav Adhikari, Michael Vilkhovoy, Sandra Vadhin, Ha Eun Lim and Jeffrey D. Varner*

Robert Frederick Smith School of Chemical and Biomolecular Engineering, College of Engineering, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Robert Parker,
University of Pittsburgh, United States

Reviewed by:

Pavel Loskot,
Swansea University, United Kingdom
Mohit Kumar Jolly,
Indian Institute of Science (IISc), India

*Correspondence:

Jeffrey D. Varner
jdv27@cornell.edu

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 28 February 2020

Accepted: 02 November 2020

Published: 26 November 2020

Citation:

Adhikari A, Vilkhovoy M, Vadhin S,
Lim HE and Varner JD (2020) Effective
Biophysical Modeling of Cell Free
Transcription and Translation
Processes.
Front. Bioeng. Biotechnol. 8:539081.
doi: 10.3389/fbioe.2020.539081

Transcription and translation are at the heart of metabolism and signal transduction. In this study, we developed an effective biophysical modeling approach to simulate transcription and translation processes. The model, composed of coupled ordinary differential equations, was tested by comparing simulations of two cell free synthetic circuits with experimental measurements generated in this study. First, we considered a simple circuit in which sigma factor 70 induced the expression of green fluorescent protein. This relatively simple case was then followed by a more complex negative feedback circuit in which two control genes were coupled to the expression of a third reporter gene, green fluorescent protein. Many of the model parameters were estimated from previous biophysical studies in the literature, while the remaining unknown model parameters for each circuit were estimated by minimizing the difference between model simulations and messenger RNA (mRNA) and protein measurements generated in this study. In particular, either parameter estimates from published studies were used directly, or characteristic values found in the literature were used to establish feasible ranges for the parameter estimation problem. In order to perform a detailed analysis of the influence of individual model parameters on the expression dynamics of each circuit, global sensitivity analysis was used. Taken together, the effective biophysical modeling approach captured the expression dynamics, including the transcription dynamics, for the two synthetic cell free circuits. While, we considered only two circuits here, this approach could potentially be extended to simulate other genetic circuits in both cell free and whole cell biomolecular applications as the equations governing the regulatory control functions are modular and easily modifiable. The model code, parameters, and analysis scripts are available for download under an MIT software license from the Varnerlab GitHub repository.

Keywords: systems biology, synthetic biological circuits, cell free, mathematical modeling, simulation

1. INTRODUCTION

Cell free systems are a widely used research tool in systems and synthetic biology and a promising platform for the manufacturing of proteins and chemicals (Vilkhovoy et al., 2020). A distinctive feature of cell free systems is the absence of cellular growth and maintenance, thereby allowing the direct allocation of carbon and energy resources toward a product of interest. Cell free systems are also more amenable than living systems to observation and manipulation, hence allowing rapid

tuning of reaction conditions. Arguably, the most widely used cell free technology is cell free protein synthesis (CFPS), an *in vitro* platform for protein transcription (TX) and translation (TL). Transcription and translation, the processes by which information stored in DNA is converted to a working protein, are at the center of metabolism and signal transduction programs important to biotechnology and human health. For example, evolutionarily conserved developmental programs such as the epithelial to mesenchymal transition (EMT) (Thiery, 2003), or retinoic acid induced differentiation (Nilsson, 1984), rely on multiple rounds of highly coordinated gene expression. From the perspective of biotechnology, even relatively simple industrially important organisms such as *Escherichia coli*, have intricate transcriptional regulatory networks which control the metabolic state of the cell in response to changing nutrient conditions (Orth et al., 2010; Vilkhovoy et al., 2016). Understanding the dynamics of regulatory networks can be greatly facilitated by mathematical models. A majority of these models fall into three categories: logical, continuous, and stochastic models (Karlebach and Shamir, 2008). Logical models such as Boolean networks (Glass and Kauffman, 1973) developed using a variety of approaches and data (Pratapa et al., 2020) represent the state of each network entity as a discrete variable, provide a quick but qualitative description of the behavior of the regulatory network. Linear and non-linear ordinary differential equation (ODE) models fall into the second category, and they generally provide a detailed picture of the network dynamics, although they can be non-physical models, e.g., relying on a gene signal perspective (Bonneau et al., 2006). Lastly, stochastic models describe the interactions between individual molecules, and discrete reaction events (McAdams and Arkin, 1997; Mettetal et al., 2006; Kaufmann and van Oudenaarden, 2007; Raj and van Oudenaarden, 2008). Model choice depends on criteria such as speed, the level of detail required and the quantity of experimental data available to estimate the model parameters. While the end goal of the models might be to accurately predict *in vivo* behavior, living systems do not necessarily provide an ideal experimental platform. For example, although there have been significant advancements in metabolomics (e.g., Park et al., 2016), the rigorous quantification of intracellular messenger RNA (mRNA) copy number or protein abundance remains challenging. Toward this challenge, cell free systems offer several advantages for the study of transcription and translation processes.

Cell free biology has historically been an important tool to study the fundamental biological mechanisms involved with gene expression. In the 1950s, cell free systems were used to explore the incorporation of amino acids into proteins (Borsook, 1950; Winnick, 1950a,b), and the role of adenosine triphosphate (ATP) in protein production (Hoagland et al., 1956). Further, *E. coli* extracts were used by Nirenberg and Matthaei in 1961 to demonstrate templated translation (Matthaei and Nirenberg, 1961; Nirenberg and Matthaei, 1961), leading to a Nobel Prize in 1968 for deciphering the codon code. More recently, as advancements in extract preparation and energy regeneration have extended their durability, the usage of cell free systems has also expanded to both small- and large-scale biotechnology and biomanufacturing applications (Swartz, 2018; Silverman et al.,

2019). Today, cell free systems have been implemented for therapeutic protein and vaccine production (Ng et al., 2012; Jaroentomeechai et al., 2018; Stark et al., 2019), biosensing (Soltani et al., 2018), genetic part prototyping (Moore et al., 2017) and minimal cell systems (Yue et al., 2019). The versatility of cell free systems offers a tremendous opportunity for the systems-level experimental and computational study of biological mechanism. For example, a number of ordinary differential equation based cell free models have been developed (Stögbauer et al., 2012; Mavelli et al., 2015; Matsuura et al., 2017; Doerr et al., 2019; Marshall and Noireaux, 2019). However, despite the obvious advantages offered by a cell free system, experimental determination of the kinetic parameters for these models is often challenging. For instance, the cell free modeling study of Horvath and coworkers (which included a description of transcription and translation, and the underlying metabolism supplying energy and precursors for transcription and translation), had over 800 unknown model parameters (Horvath et al., 2020). Moreover, the construction, identification and validation of the Horvath model took well over a year to complete. Thus, constructing, identifying and validating biophysically motivated cell free models, which are simultaneously manageable, is a key challenge. Toward this challenge, effective modeling approaches which coarse grain biological details but remain firmly rooted in a biophysical perspective, could be an important tool.

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

either parameter estimates from published studies were used directly, or characteristic values found in the literature were used to establish feasible ranges for the parameter estimation problem. Next, in order to provide a detailed insight into the influence of individual model parameters on the expression dynamics of each circuit, Morris sensitivity analysis was employed. For $C1$, the sensitivity results were informative, but expected. However, for $C2$, the analysis hierarchically stratified the parameters (and associated model species) into local vs. global categories. For example, parameters that controlled the abundance of lambda phage repressor protein (cI-ssrA), a master circuit regulator in $C2$, were globally important as they influenced all other species. On the other hand, the parameters that influenced deGFP levels (the endpoint of both circuits) were only locally important to deGFP. Taken together, the effective biophysical modeling approach captured the expression dynamics, including the transcription dynamics, for two synthetic cell free circuits. While, we considered only two circuits here, this approach could potentially be extended to simulate other genetic circuits in both cell free and whole cell biomolecular applications. The model code, parameters, and analysis scripts are available under an MIT software license from the Varnerlab GitHub repository¹.

2. MATERIALS AND METHODS

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 μ L, composed of the Sigma 70 Master Mix (9 μ L) and the plasmids (3 μ L total): P70a-deGFP (5 nM) for the single-gene system; P70a-deGFP-ssrA (8 nM), P70a-S28 (1.5 nM), and P28a-cI-ssrA (1 nM) for the negative feedback circuit. The plasmids were bought in lyophilized form (Arbor Biosciences) and purified using QIAprep Spin Miniprep Kit (Qiagen) using cell lines DH5-Alpha (for P28a-cI-ssrA) or KL740 (for P70a-deGFP, P70a-deGFP-ssrA, and P70a-S28). The CFPS reactions were incubated at 29°C.

2.2. mRNA and Protein Quantification

Following each CFPS run, the total RNA was extracted from 1 μ L of the reaction mixture using PureLink RNA Mini Kit (Thermo Fisher Scientific) and stored at -80°C. The quantitative RT-PCR reactions were done using Applied Biosystems™ TaqMan™ RNA-to-CT™ 1-Step Kit and Custom TaqMan Gene Expression Assays (Thermo Fisher Scientific). An mRNA standard curve was used to determine absolute mRNA concentrations for each of the samples. The mRNA standards were prepared as follows: separate CFPS reactions for 5 nM of plasmids (P70a-S28, P70a-deGFP, and P70a-deGFP-ssrA) were carried out for 2 h. Total RNA was extracted using the full reaction volume. This was followed by the removal of 16S and 23S rRNA using the MICROExpress™ Bacterial mRNA Enrichment Kit (Life Technologies Corporation). Lastly, the MEGAclean™ Kit (Life Technologies Corporation) was used

to further purify the mRNA. The mRNA concentrations were determined using the Qubit™ RNA assay kit (ThermoFisher Scientific). At least three technical replicates were performed for each standard. The concentration of cI-ssrA mRNA was quantified using the deGFP-ssrA mRNA standard. Green fluorescent protein (deGFP) fluorescence was measured using the Varioskan Lux plate reader at 488 nm (excitation) and 535 nm (emission). At the end of the CFPS run, 3 μ L of the reaction mixture was diluted in 33 μ L phosphate buffered saline (PBS) and stored at -80°C. The fluorescence was measured in triplicate with 10 μ L each of this mixture. For all measurements, at least three biological replicates were performed.

2.3. Synthetic Circuit Architecture

The two genetic circuits ($C1$ and $C2$) used in this study were based upon the bacterial sigma factor regulatory system (Figure 1). Sigma factor 70 (σ_{70}), endogenously present in the extract, was the primary driver of each circuit. In $C1$, σ_{70} induced green fluorescent protein (deGFP) expression was explored in the absence of additional regulators or protein degradation (Figure 1A). In $C2$, σ_{70} induced the expression of sigma factor 28 (σ_{28}) and deGFP-ssrA (Figure 1B). Sigma 28 induced the expression of the lambda phage repressor protein cI-ssrA, which was under the σ_{28} responsive P28 promoter. The cI-ssrA protein repressed the P70a promoter, thereby down-regulating σ_{28} and deGFP-ssrA transcription (Marshall and Noireaux, 2018). Simultaneously, the C-terminal ssrA degradation tags present on the deGFP and cI proteins were recognized by the endogenous ClpXP protease system in the cell free extract, thereby promoting the degradation of these proteins into peptide fragments (Flynn et al., 2003; Garamella et al., 2016). In addition, messenger RNAs (mRNAs) were always subject to degradation due to the presence of degradation enzymes in the extract (Karzbrun et al., 2011; Garamella et al., 2016). Taken together, the interactions of the components manifested in an accumulation of deGFP protein for $C1$, and a pulse signal of deGFP-ssrA in $C2$. Studying $C1$ allowed us to estimate parameters governing the interaction of σ_{70} with the P70a promoter. Whereas, the $C2$ allowed us to characterize the interaction of σ_{28} with the P28 promoter, the strength of the transcriptional repression by cI-ssrA, and the kinetics of protein degradation by the endogenous ClpXP protease system. Finally, both circuits tested the effective model formulation for the transcription and translation rates.

2.4. Formulation and Solution of Model Equations

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

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

$$\dot{p}_j = r_{L,j}w_j(\dots) - \theta_{p,j}p_j \quad (2)$$

The term $r_{X,j}u_j(\dots)$ in the mRNA balance, which denotes the regulated rate of transcription for gene j , is the product of a kinetic limit $r_{X,j}$ (nM h^{-1}) and a transcription control function $0 \leq u_j(\dots) \leq 1$ (dimensionless). Similarly, the rate of translation

¹Varnerlab. Github repository for tx/tl model code. Available online at <https://github.com/varnerlab/Biophysical-TXTL-Model-Code>.

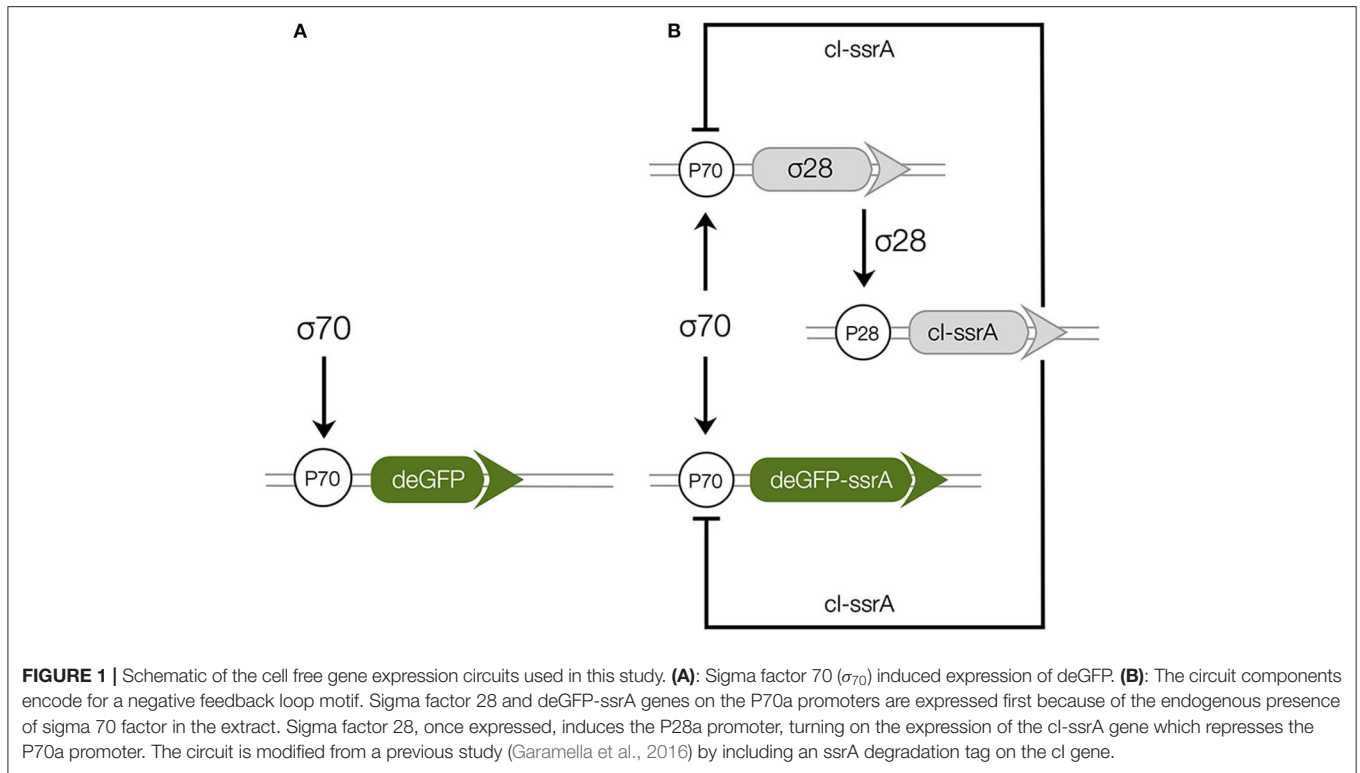


FIGURE 1 | Schematic of the cell free gene expression circuits used in this study. **(A)**: Sigma factor 70 (σ_{70}) induced expression of deGFP. **(B)**: The circuit components encode for a negative feedback loop motif. Sigma factor 28 and deGFP-ssrA genes on the P70a promoters are expressed first because of the endogenous presence of sigma 70 factor in the extract. Sigma factor 28, once expressed, induces the P28a promoter, turning on the expression of the cl-ssrA gene which represses the P70a promoter. The circuit is modified from a previous study (Garamella et al., 2016) by including an ssrA degradation tag on the cl gene.

of mRNA j , denoted by $r_{L,j}w_j$, is also the product of the kinetic limit of translation (nM h^{-1}) and a translational control term $0 \leq w_j(\dots) \leq 1$ (dimensionless). Lastly, $\theta_{*,j}$ denotes the first-order rate constant (h^{-1}) governing degradation of protein and mRNA. The model equations, encoded in the Julia programming language (Bezanson et al., 2017), were automatically generated using the JuGRN tool². The model equations were solved numerically using the Rosenbrock23 routine of the DifferentialEquations.jl package (Rackauckas and Nie, 2017).

2.4.1. Transcription and Translation Kinetic Limits

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

$$r_{X,j} = V_{X,j}^{max} \left(\frac{G_j}{\tau_{X,j}K_{X,j} + (1 + \tau_{X,j})G_j + \mathcal{O}_{X,j}} \right) \quad (3)$$

²Varnerlab. Gene Regulatory Network Generator in Julia (JuGRN). Available online at <https://github.com/varnerlab/JuGRN-Generator>.

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

$$\mathcal{O}_{X,j} = \sum_{i=1,j}^N \frac{K_{X,j}\tau_{X,j}}{K_{X,i}\tau_{X,i}} (1 + \tau_{X,i}) G_i \quad (4)$$

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

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

$$r_{L,j} = V_{L,j}^{max} \left(\frac{m_j}{\tau_{L,j}K_{L,j} + (1 + \tau_{L,j})m_j + \mathcal{O}_{L,j}} \right) \quad (5)$$

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

$$\mathcal{O}_{L,j} = \sum_{i=1,j}^N \frac{K_{L,j}\tau_{L,j}}{K_{L,i}\tau_{L,i}} (1 + \tau_{L,i}) m_i \quad (6)$$

describes the coupling of the translation of mRNA j with other messages in the system because of kinetic competition for available ribosomes. The saturation and time constants for

each case (which are unknown and must be estimated from experimental data) are defined in the **Supplementary Materials**. Lastly, in this study, the $\mathcal{O}_{X,j}$ and $\mathcal{O}_{L,j}$ terms were neglected as both circuits had either only one, or a small number of genes.

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

$$V_{X,j}^{max} \equiv R_{X,T} \left(\frac{\dot{v}_X}{l_{G,j}} \right) \quad (7)$$

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

$$V_{L,j}^{max} \equiv K_P R_{L,T} \left(\frac{\dot{v}_L}{l_{P,j}} \right) \quad (8)$$

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

2.4.2. Control Functions u and w

Values of the control functions $u(\dots)$ and $w(\dots)$ describe the regulation of transcription and translation. Ackers et al., borrowed from statistical mechanics to recast the $u(\dots)$ function as the probability that a system exists in a configuration which leads to expression (Ackers et al., 1982). The idea of recasting $u(\dots)$ as the probability of expression was also developed (apparently independently) by Bailey and coworkers in a series of papers modeling the *lac* operon (see Lee and Bailey, 1984). More recently, Moon and Voigt adapted a similar approach when modeling the expression of synthetic circuits in *E. coli* (Moon et al., 2012). The $u(\dots)$ function is formulated as:

$$u(\dots)_j = \frac{\sum_{i \in \{\chi\}} W_i f_i(\dots)}{\sum_{j \in \{C_j\}} W_j f_j(\dots)} \quad (9)$$

where W_i (dimensionless) denotes the weight of configuration i , while $f_i(\dots)$ (dimensionless) is a binding function (taken to be a hill-type function) describing the fraction of bound activator/inhibitor for configuration i . The summation in the numerator of Equation (9) is over the set of promoter configurations leading to expression (denoted as χ), while the summation in the denominator is over the set of all possible configurations for gene j (denoted as C_j). Thus, $u(\dots)_j$ can be thought of as the fraction of all possible configurations that lead to expression. The weights W_i are related to the Gibbs energy of configuration i : $W_i = \exp(-\Delta G_i/RT)$ where ΔG_i denotes the molar Gibbs energy for configuration i (kJ/mol), R denotes the ideal gas constant ($\text{kJ mol}^{-1} \text{K}^{-1}$), and T denotes the system temperature (Kelvin) (Ackers et al., 1982). The value of the binding function depends on the concentrations of the different transcriptional elements and their dissociation constants. The temporal evolution of u , therefore, is tied to the dynamics of its

transcriptional elements, and its value lies between 0 and 1. In the case of circuit $C1$, u did not vary during the course of the reaction because the concentration of its activator, σ_{70} , was fixed. For this case, u approximately equalled 0.95. However, in the second circuit, $C2$, u varied with time because of the change in levels of σ_{28} and cI-ssrA proteins.

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

$$\dot{\epsilon} = - \left(\frac{0.693}{\tau_{L,1/2}} \right) \epsilon \quad (10)$$

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

2.5. Estimation of Model Parameters

Model parameters were estimated from published studies, were specified by experimental conditions (Table 1) or were estimated by minimizing the squared difference between model simulations and messenger RNA (mRNA), or protein measurements generated in this study. For the P70-deGFP model ($C1$), 11 parameters were estimated, while 33 parameters were estimated for the negative feedback model ($C2$).

The minimization problem to estimate the unknown model parameters was structured as a multiobjective optimization problem in which each measured mRNA or protein trajectory was treated as a separate objective. The minimization problem was solved using the Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) (Bassen et al., 2017). JuPOETs is a multiobjective optimization approach which integrates simulated annealing with Pareto optimality to estimate parameter values on or near the optimal tradeoff surface between N potentially competing objectives (squared difference between model simulations and experimental measurements). JuPOETs minimized a problem of the form:

$$\min_k \mathcal{E}_j = \sum_{i=1}^{\mathcal{T}_j} \left(\hat{\mathcal{M}}_{ij} - x_{ij}(k) \right)^2 \quad j = 1, 2, \dots, N \quad (11)$$

subject to

$$\dot{x} = f(x, k) \quad (12)$$

$$\mathcal{L} \leq k \leq \mathcal{U} \quad (13)$$

$$x(t_0) = x_0 \quad (14)$$

where Equation (12) denotes the model equations, Equation (13) denotes the parameter bounds, and Equation (14) denotes

TABLE 1 | Characteristic parameters for TX/TL model equations.

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

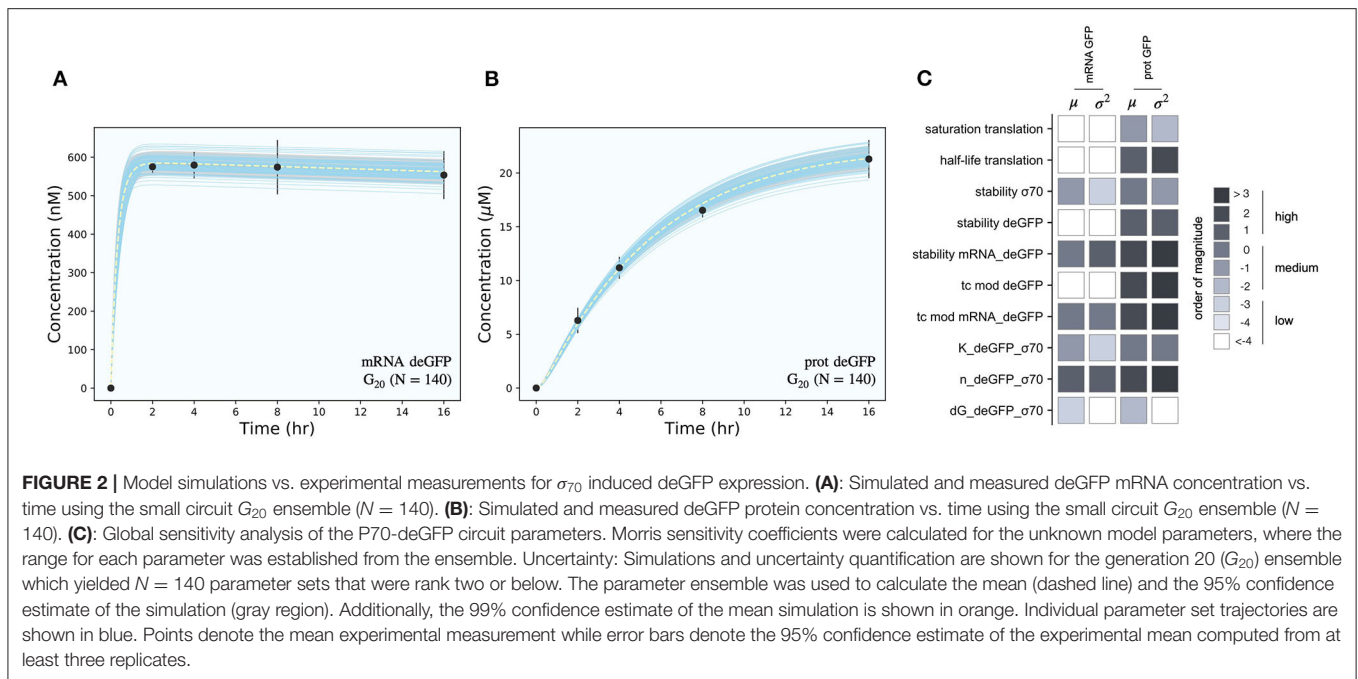
Key to references used in the table: (a) Garamella et al. (2016), (b) Underwood et al. (2005), (c) set by experiment, (d) Kassavetis and Chamberlin (1981), (e) estimated in this study, (f) Niederholtmeyer et al. (2015) and Vilkhovoy et al. (2018), (g) Grilly et al. (2007), (h) calculated from plasmid sequence, (i) McClure (1980).

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

The lower and upper bounds for unknown model parameters were established from previously published studies, or from previous model analysis; parameter values estimated for the P70-deGFP model were also used to establish ranges for the negative feedback model. JuPOETs searched over ΔG_i , K_L , and $\tau_{L,1/2}$ values directly, while other unknown parameter values took the form of corrections to order of magnitude characteristic literature estimates. For example, we set the mRNA degradation

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

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



low rank low error solutions is slow. To address this concern, we periodically switch to single objective mode where we minimize the total training error (summation of all objective functions) instead of finding low rank solutions. The best solutions from single objective mode can then be used to restart the multi-objective calculation. This hybrid approach, which was used in this study, has previously been shown to increase the rate of finding low rank and low error solutions (see Bassen et al., 2017).

2.6. Morris Sensitivity Analysis

Morris sensitivity analysis was used to understand which model parameters were sensitive (Morris, 1991). The Morris method is a global method that computes an elementary effect value for each parameter by sampling a model performance function, in this case the area under the curve for each model species in their respective timeplots, over a range of values for each parameter; the mean of elementary effects measures the direct effect of a particular parameter on the performance function, while the variance of each elementary effect indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests connectedness or non-linearity). The Morris sensitivity coefficients were computed using the `DiffEqSensitivity.jl` package (Rackauckas and Nie, 2017). The parameter ranges were established by calculating the minimum and the maximum value for each parameter in the parameter ensemble generated by JuPOETs. Each range was then subdivided into 10,000 samples for the sensitivity calculation. Elementary effect values were then calculated one at a time by measuring the perturbation in the AUC on changing one parameter, where the AUC was calculated by solving the set of ODEs for each change. In order to calculate the mean and variance, the top 1000 perturbations with the

highest spread in parameter values were used. In total, the model was evaluated 100000n times, where n is the number of parameters varied. The Morris sensitivity coefficients are calculated using the `compute_sensitivity_coefficients.jl` script in the model repositories.

3. RESULTS

3.1. Modeling and Analysis of the C1 Circuit

The effective biophysical transcription and translation model captured σ_{70} induced deGFP expression at the mRNA and protein level within the experimental error for C1 (Figure 2). JuPOETs produced an ensemble ($N = 140$) of the 11 unknown model parameters which captured the transcription of mRNA (Figure 2A) and the translation of deGFP protein (Figure 2B). The mean and standard deviation of key parameters is given in Table 2. The deGFP mRNA reached its steady state concentration of approximately 580 nM within 2 h, and stayed at this level for the remainder of the reaction. Thus, the cell free reaction maintained continuous transcriptional activity with an average mRNA lifetime of 27 min; Garamella et al. (2016) reported a similar lifetime of 17–18 min. On the other hand, deGFP protein concentration increased more slowly, and began to saturate between 8 and 10 h at approximately 15 μ M. Given there was negligible protein degradation (the mean deGFP half-life was estimated to be \sim 11 days, which was similar to the value of 6 days estimated by Horvath et al., albeit in a different cell free system, Horvath et al., 2020). The saturating protein concentration suggested that the translational capacity of the cell free system decreased over the course of the reaction. The decrease in translational capacity, which could stem from several sources, was captured in the simulations using a monotonically decreasing

TABLE 2 | Estimated parameter values for the P70-deGFP model (C1).

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

The mean and standard deviation of each parameter value was calculated over the ensemble of parameter sets meeting the rank selection criteria ($N = 139$).

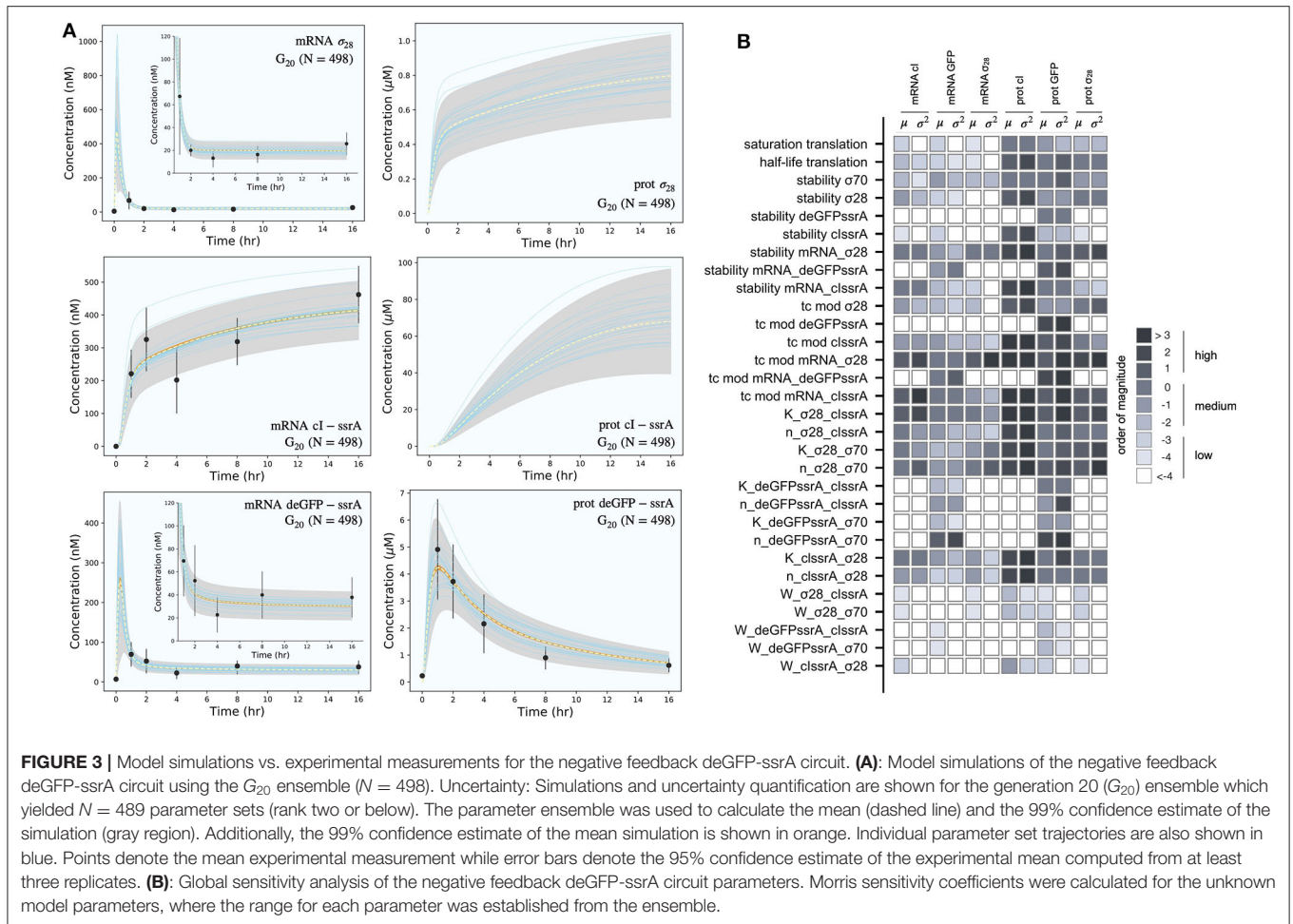
translation capacity state variable ϵ , and the translational control variable $w(\dots)$. In particular, the mean half-life of translational capacity was estimated to be $\tau_{L,1/2} \sim 4$ h in the C1 experiments. Taken together, JuPOETs produced an ensemble of model parameters that captured the experimental training data. Next, we considered which C1 model parameters were important to the model performance using Morris sensitivity analysis, a global sensitivity analysis method.

The importance of C1 model parameters was quantified using Morris sensitivity analysis (Figure 2B). The Morris method computes the influence of each parameter, known as the elementary effect, on a model performance function. The mean of elementary effects measures the direct effect of a particular parameter, while the variance indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests non-linearity). The performance function was defined as the integrated area under the curve (AUC) for each mRNA and protein species in their respective timeplots, calculated for each parameter value range. The Morris sensitivity measures (mean and variance) were binned into categories based upon their relative magnitudes, from no influence (white) to high influence (black). Only four parameters (translation saturation coefficient K_L , translational capacity half-life $\tau_{L,1/2}$, translation time constant, and protein degradation constant) influenced the protein level. On the other hand, six parameters influenced both mRNA and protein abundance; all six of these parameters were either directly or indirectly associated with transcription. Thus, these parameters influenced the production or stability of mRNA which in turn influenced the protein level. In particular, the mRNA degradation constant and the cooperativity of σ_{70} in the P70a promoter function had the largest direct effect and variance. Surprisingly, the ΔG of σ_{70} /RNAP/promoter configuration was the least influential of the six parameters and had a small elementary effect variance. Taken together, Morris sensitivity analysis of

the C1 model parameters highlighted the hierarchical structure of the transcriptional and translational model, suggesting experimentally tunable parameters such as mRNA stability were globally important. Next, we used the ensemble of P70a, time constant and degradation parameters estimated for C1 to constrain the analysis of C2.

3.2. Modeling and Analysis of the C2 Circuit

The effective biophysical transcription and translation model captured the deGFP-ssrA expression dynamics in the negative feedback circuit C2 (Figure 3A). JuPOETs produced an ensemble ($N = 498$) of the 33 unknown model parameters which captured transcription and translation dynamics for σ_{28} , cI-ssrA and deGFP-ssrA. The mean and standard deviation of key parameters is given in Table 3. Compared with the estimated parameters for C1, the C2 model had almost a two fold change in the half life of translation and the translation saturation coefficient. Similarly, there were variations in the values of the transcription and translation time constants for the two systems. However, for both circuits, the small values of the transcription and translation time constants qualitatively suggested elongation limited reactions; the exception was σ_{28} translation which was closer to initiation limited. Unlike C1, the mRNA expression pattern for σ_{28} and deGFP-ssrA both showed an initial spike, to a concentration similar with the previous pseudo steady state, before the cI-ssrA regulator protein could be expressed. However, once cI-ssrA began to accumulate, the concentrations of the regulated mRNAs dropped by approximately an order of magnitude compared with the unregulated case. Again, as shown with C1, the regulated mRNA concentrations reached an approximate steady-state. This further confirmed continuous transcription and mRNA degradation throughout the cell free reaction. The mean estimated mRNA lifetime for cI-ssrA and deGFP were similar (approximately 16 min), while the degradation of σ_{28} mRNA was predicted to be slower (mean mRNA lifetime



was estimated to be approximately 30 min). Lastly, the mean peak degradation rate for GFP was approximately 47 nM/min, while the mean peak cI-ssrA degradation rate was predicted to be approximately 63 nM/min; both of these degradation rate estimates were consistent with the previously reported range of 15–150 nM/min measured by Garamella et al. (2016).

The secondary effect of cI-ssrA repression was visible in the cI-ssrA mRNA expression pattern. The expression of cI-ssrA was induced by σ_{28} , however, σ_{28} expression was repressed by cI-ssrA, thereby completing a negative feedback loop. Initially, before appreciable levels of cI-ssrA had been translated, the cI-ssrA transcription rate was maximum (approximately 200 nM/h). However, the transcription rate decreased to approximately 12 nM/h after 2 h and remained constant for the remainder of the cell free reaction. Similarly, transcription rates for σ_{28} (approximately 1,200 nM/h) and deGFP-ssrA (approximately 750 nM/h) were initially at a maximum due to the presence of endogenous σ_{70} , but then quickly dropped as cI protein accumulated. Protein synthesis followed a similar trend, with the translation rates for σ_{28} and deGFP-ssrA initially present at their maximum values before quickly dropping. After 1 h, deGFP levels reached a peak and decayed due to the ClpXP-mediated degradation, whereas σ_{28} protein levels continued to slowly rise at a steady rate (approximately 15 nM/h). The $C2$

model also predicted the expected lag present during the initial phase of cI-ssrA protein synthesis due to the need for σ_{28} protein to reach appreciable levels. Moreover, the combination of high cI-ssrA mRNA abundance (expressed because σ_{28} does not have a degradation tag) and ClpXP-mediated degradation led to the saturation of the cI-ssrA protein concentration. However, the cI-ssrA protein concentration could not be verified because we did not have an experimental measurement for this species. Taken together, the effective model simulated cell free expression dynamics for $C2$. Next, we considered which $C2$ model parameters were important using Morris sensitivity analysis.

Morris sensitivity analysis of the negative feedback circuit $C2$ stratified the parameters into locally and globally important groups (Figure 3B). The influence of 33 parameters was computed using the AUC of each mRNA and protein species as the performance function. The Morris sensitivity metrics (mean and variance) were binned into categories based upon their relative magnitudes, from no influence (white) to high influence (black). Some parameters affected only their respective mRNA or protein target, whereas others had widespread effects. For example, the time constant (tc) modifiers, stability of deGFP-ssrA protein and mRNA, and the binding dissociation constant (K) and cooperativity parameter (n) of cI-ssrA and σ_{70} for the deGFP-ssrA promoter affected only the values of deGFP-ssrA

TABLE 3 | Estimated parameter values for the negative feedback circuit (C2).

Description	Parameter	Value ($\mu \pm \sigma$)	Units
Translation saturation coefficient	K_L	253.75 ± 14.12	μM
Half-life translation	$\tau_{L,1/2}$	8.86 ± 0.85	h^{-1}
Time constants			
cl-ssrA transcription	$\tau_{X,cl}$	< 0.001	dimensionless
deGFP transcription	$\tau_{X,GFP}$	0.045 ± 0.003	dimensionless
σ_{28} transcription	$\tau_{X,\sigma_{28}}$	0.0018 ± 0.0003	dimensionless
cl-ssrA translation	$\tau_{L,cl}$	0.054 ± 0.004	dimensionless
deGFP translation	$\tau_{L,GFP}$	0.058 ± 0.007	dimensionless
σ_{28} translation	$\tau_{L,\sigma_{28}}$	1.1 ± 0.13	dimensionless
mRNA and protein half-life			
mRNA cl-ssrA	$\ln(2)/\theta_{m,cl}$	8.1 ± 0.60	min
mRNA deGFP	$\ln(2)/\theta_{m,GFP}$	7.74 ± 1.13	min
mRNA σ_{28}	$\ln(2)/\theta_{m,\sigma_{28}}$	14.96 ± 1.60	min
Protein cl-ssrA	$\ln(2)/\theta_{p,cl}$	0.46 ± 0.043	days
Protein deGFP-ssrA	$\ln(2)/\theta_{p,GFP}$	0.051 ± 0.002	days
Protein σ_{28}	$\ln(2)/\theta_{p,\sigma_{28}}$	7.65 ± 0.91	days
Protein σ_{70}	$\ln(2)/\theta_{p,\sigma_{70}}$	14.86 ± 2.30	days
Free energies			
RNAP + cl gene	$\Delta G_{cl,RX}$	46.57 ± 4.28	kJ mol^{-1}
RNAP + σ_{28} + cl gene	$\Delta G_{cl,\sigma_{28}}$	-1.10 ± 0.04	J mol^{-1}
RNAP + deGFP gene	$\Delta G_{GFP,RX}$	41.94 ± 1.80	kJ mol^{-1}
RNAP + σ_{70} + deGFP gene	$\Delta G_{GFP,\sigma_{70}}$	-27.67 ± 1.79	kJ mol^{-1}
RNAP + cl + deGFP gene	$\Delta G_{GFP,cl}$	-7.21 ± 1.14	kJ mol^{-1}
RNAP + σ_{28} gene	$\Delta G_{\sigma_{28},RX}$	46.67 ± 3.18	kJ mol^{-1}
RNAP + σ_{70} + σ_{28} gene	$\Delta G_{\sigma_{28},\sigma_{70}}$	-10.46 ± 1.15	kJ mol^{-1}
RNAP + cl + σ_{28} gene	$\Delta G_{\sigma_{28},cl}$	-12.89 ± 1.44	kJ mol^{-1}
Hill coefficients			
cl gene + σ_{28}	$n_{cl,\sigma_{28}}$	1.88 ± 0.28	dimensionless
deGFP gene + σ_{70}	$n_{GFP,\sigma_{70}}$	1.53 ± 0.14	dimensionless
deGFP gene + cl	$n_{GFP,cl}$	0.698 ± 0.133	dimensionless
σ_{28} gene + σ_{70}	$n_{\sigma_{28},\sigma_{70}}$	1.10 ± 0.10	dimensionless
σ_{28} gene + cl	$n_{\sigma_{28},cl}$	1.51 ± 0.25	dimensionless
Dissociation constants			
cl gene + σ_{28}	$K_{cl,\sigma_{28}}$	1.09 ± 0.088	μM
deGFP gene + σ_{70}	$K_{GFP,\sigma_{70}}$	86.87 ± 7.13	μM
deGFP gene + cl	$K_{GFP,cl}$	3.83 ± 0.41	μM
σ_{28} gene + σ_{70}	$K_{\sigma_{28},\sigma_{70}}$	1.35 ± 0.26	μM
σ_{28} gene + cl	$K_{\sigma_{28},cl}$	0.0389 ± 0.0068	μM

The mean and standard deviation for each parameter was calculated over the ensemble of parameter sets ($N = 498$).

protein and mRNA. On the other hand, the tc, stability, K and n parameters for σ_{70} , σ_{28} , or cI-ssrA influenced mRNA and protein expression globally. The σ_{70} and σ_{28} proteins acted as inducers or repressors for more than one gene product: σ_{70} induced both deGFP-ssrA and σ_{28} , and cI-ssrA protein repressed both of these genes. Degradation constants (denoted as stability) affected the half-lives of the transcribed messages or the translated proteins in the mixture, while the time constant modifiers changed the time required to form the open gene complex (or translationally active complex). Dissociation and cooperativity constants affected the binding interactions of the inducer (or repressor in the case of cI-ssrA) in the promoter control function. Varying these parameters, therefore, had a strong effect on their respective targets. Similarly, the translation saturation and its half-life,

which captured the depletion in the translation activity over the course of the reaction, not only affected protein levels but also mRNA levels. This is because these parameters tuned the rate of formation of cI-ssrA, which in turn affected the mRNA levels of its gene targets. Given that cI-ssrA was the main regulator (repressor) of the circuit, the parameters that dictated the levels of cI-ssrA mRNA and protein had a global effect. We also observed high sensitivity variance for several parameters, in particular involving cI-ssrA. For example, the time constant modifiers for cI-ssrA mRNA and protein had a two-pronged effect. On the one hand, they positively influenced the transcription/translation rates of the gene and mRNA products, directly increasing the cI-ssrA protein. On the other hand, increased cI-ssrA expression reduced the level of σ_{28} , in turn reducing the cI-ssrA levels

over time. Taken together, Morris sensitivity analysis of the $C2$ model stratified that parameters into local and globally important groups, with the parameters governing the synthesis rates of the *cI*-*ssrA* mRNA and protein being globally important. The sensitivity analysis also gave insight into the organization of the circuit, suggesting *cI* to be highly connected within the circuit.

4. DISCUSSION

In this study, we developed an effective biophysical modeling approach to simulate transcription (TX) and translation (TL) processes occurring in a cell free system. We tested this approach by simulating the dynamics of two cell free synthetic circuits ($C1$ and $C2$).

The model formulation, and parameter values were mechanistic and largely derived from literature. For example, characteristic values for τ_X and K_X , the time and saturation constants for transcription, were approximated from *in vitro* experiments using an abortive initiation assay (McClure, 1980). The RNAP and ribosome elongation rates were taken from Garamella et al. (2016), while other parameters were estimated from BioNumbers (Milo et al., 2010). Similarly, the weights appearing in the transcription control function $u(\dots)$ were based upon the Gibbs energies of the respective promoter configurations, while the form of the transcriptional control functions was derived from a statistical mechanical treatment of promoter activity (Ackers et al., 1982; Lee and Bailey, 1984; Moon et al., 2012). However, there were parameters that were not available from literature; in these cases multiobjective optimization was used to estimate these parameters from mRNA and protein measurements. For $C1$, sigma factor 70 (σ_{70}) induced expression of green fluorescent protein (deGFP), the time constants, degradation rates, and other parameters governing deGFP expression were estimated from measurements of deGFP mRNA and protein. These estimates were then used to constrain the parameter search for $C2$, which involved deGFP expression subject to negative feedback and programmed protein degradation. We estimated which model parameters were important to the performance of $C1$ and $C2$ using Morris sensitivity analysis. Sensitivity analysis results for $C1$ were expected; the time constant for transcription, the stability of the deGFP message and the cooperativity of σ_{70} were all important parameters. On the other hand, the sensitivity analysis results for $C2$ were more nuanced, with parameters (and associated species) being stratified into locally and globally important groups; the performance of $C2$ was most sensitive to the parameters controlling the *cI*-*ssrA* mRNA and protein abundance.

The effective TX/TL modeling approach described here has several potential applications. For example, a challenge of *in vivo* constraint based modeling is the description of gene expression (Covert and Palsson, 2002). Boolean and probabilistic approaches (Covert et al., 2001, 2004; Chandrasekaran and Price, 2010) have been developed to address this challenge. However, the transcriptional state of a boolean gene is either on or off based on the state of its regulators, thus, gene expression is coarse-grained. The current modeling approach could be an interesting mechanistic alternative to the boolean framework that utilizes a continuous description of gene expression dynamics

and transcriptional regulation. In particular, the rules encoding typical boolean gene expression networks are easily translatable into the rational promoter functions described here, however, the estimation of the parameters appearing in these promoter functions, especially in an *in vivo* context, remains an open question. Another application could be the extension of the current model to other prokaryotic or eukaryotic systems with a few changes. For example, in order to adopt it for an *in vivo* system, the dilution of resources due to growth (proportional to the cellular doubling time) would be added as a first order term to the mRNA and protein balance equations. Additionally, the competition for RNAP and ribosomes, denoted respectively as $O_{X,j}$ and $O_{L,j}$ in the study, and assumed to be negligible due to the presence of only three genes in the system, would need to be taken into account; this term would serve to change the rates of transcription and translation of the added genes because of the presence of a large amount of endogenous genes in the *in vivo* system. Moreover, characteristic literature-based parameter values would be different for cellular processes compared to the *in vitro* ones used in this study, and they would thus need to be adjusted accordingly. For the case of a mammalian or a yeast *in vivo* system, a few more changes to the current model are necessary because the mechanistic processes of gene expression and regulation are different in these two types of systems. For example, a key difference present in eukaryotes is the addition of an intron splicing step during the synthesis of a mature mRNA from a pre-mRNA. In addition, the gene regulation mechanisms are vast and composed of numerous elements in eukaryotes. Finally, especially in *in vivo* systems, addition of exogenous genes often leads to a tug-of-war of carbon and energy resources between cellular growth processes and the expression of these genes, driving cellular resources away from the latter. Synthetic biology studies often neglect the role that metabolism plays in the expression of synthetic circuits. Ultimately, metabolism is centrally important to the operation of any synthetic circuit; gene expression is strongly dependent upon the metabolic resources provided by catabolic processes. It is imperative that this metabolic burden by the addition of exogenous genes be incorporated in the *in vivo* model description to accurately capture the expression behavior. We have recently started to explore this question by integrating effective transcription and translation models with metabolic networks in cell free reactions e.g., Vilkhovoy et al., 2018; Horvath et al., 2020, and also developing experimental tools to measure metabolite concentrations in cell free systems (Vilkhovoy et al., 2019). However, these previous transcriptional and translational models (and similar precursor models simulating eukaryotic processes, Gould et al., 2016; Tasseff et al., 2017) were less developed than those presented here. Taken together, the effective modeling approach described here can potentially be used to simulate transcription and translation processes in a variety of applications.

There have been many studies looking into oscillatory and other dynamic behavior of synthetic circuits (see Prangemeier et al., 2020). A negative feedback loop, such as the one explored here, has the potential to give rise to oscillations. Yelleswarapu et al. carried out TX/TL reactions, with a circuit similar to $C2$, in both batch and continuous conditions (Yelleswarapu et al., 2018).

Similar to our study, no oscillations were observed in the batch reaction. However, oscillations were observed in the continuous reaction. There are several possible reasons why no oscillations were seen in our (or the Yelleswarapu et al.) batch study; as it was carried out in batch, dilution of the expressed protein or mRNA species due to an inlet feed was not possible. Thus, mRNA species reached a pseudo steady state (after approximately 2 h) because of ribonuclease degradation (Garenne et al., 2019). On the other hand, in general protein species were not at steady-state; only proteins tagged with a *ssrA* tag were able to be degraded by the ClpXP system, thereby allowing a steady-state. Thus, the batch system likely evolved dynamically through a set of concentration profiles that were not consistent with oscillations.

The effective TX/TL model described the experimental mRNA and protein training data. However, there were several important questions to be addressed by future studies. First, the model formulation described the data, but did not predict dynamics outside of the training set. If this approach is to be useful to the synthetic biology community, or more broadly as an effective biophysical technique to model *in vivo* gene expression dynamics for applications such as regulatory flux balance analysis, we need to have confidence that the modeling approach is predictive. Thus, while we have established a descriptive model, we have yet to establish a predictive model. Next, there were several technical or mechanistic questions that should be explored further. For example, *cI-ssrA* represses the activity of the P70a promoter via interaction with its OR2 and OR1 operator sites; in this study we considered only a single operator site suggesting that we potentially underestimated the potency of *cI* repression in the deGFP and σ_{28} promoter functions, see the multiplication rule (Lucks et al., 2011). Further, we used a first order approximation of ClpXP mediated protein degradation, while Garamella et al. (2016) described this degradation as zero order. Similarly, we did not establish the concentration of ClpXP in the commercially available cell free reaction mixture. The levels of this protein complex could be an important factor controlling protein degradation. Next, we should compare the current modeling approach, and the values estimated for the model parameters, with the study of Marshall and Noireaux (2019). For example, one of the potential limitations of the current study (that was addressed by Marshall and Noireaux, 2019) is that we did not consider a separate species for dark GFP. In our previous RNA circuit modeling (Hu et al., 2015), we did include this term, but failed to do so here. We expect inclusion of a dark vs. light GFP species could influence the values for the estimated parameters, particularly the translation time constants. However, previous reports suggested the *in vitro* maturation time of deGFP was approximately 8 min (Shin and Noireaux, 2010), much faster than the typical maturation times for GFP of 1 h *in vivo* (Sniegowski et al., 2005; Iizuka et al., 2011). Thus, the impact of including a dark vs. light GFP species may not be worth the increased model complexity. Lastly, we should validate the values estimated for the binding function parameters and the promoter configuration free energies. Maeda et al. measured the binding affinities of the seven *E. coli* σ factors with RNAP (Maeda et al., 2000); while not directly comparable, these measurements give an order of magnitude characteristic value for the dissociation

constants appearing in the promoter binding functions. Further, there have been several studies that have quantified the binding energies of promoter configurations (e.g., Ackers et al., 1982; Brewster et al., 2012; Tapia-Rojo et al., 2012, 2014). A perfunctory inspection of the values estimated in this study suggested our estimated free energy values, while the same order of magnitude as previous studies in many cases, did have values that were off by a factor of up to an order of magnitude compared with literature (albeit for different promoters). In particular, the positive Gibbs energy estimated for free RNAP binding leading to transcription was likely too large, while the magnitude of other values such as the energy of *cI* repression of σ_{28} expression was likely too small. Thus, these other studies could serve as a basis to validate our estimates, and perhaps more importantly constrain the parameter search space for future studies.

DATA AVAILABILITY STATEMENT

Model code is available under an MIT software license from the Varnerlab GitHub repository for TX/TL model code (<https://github.com/varnerlab/Biophysical-TX/TL-Model-Code>). The mRNA and protein measurements presented in this study are available in the data directory of the model repositories in comma separated value (CSV) and Microsoft Excel format.

AUTHOR CONTRIBUTIONS

JV directed the study. MV, SV, HL, and AA conducted the cell free experimental measurements. JV, MV, and AA developed the reduced order models and the parameter ensemble. MV, AA, and JV analyzed the model ensemble, and generated figures for the manuscript. The manuscript was prepared and edited for publication by AA, MV, SV, and JV. All authors reviewed this manuscript.

FUNDING

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

ACKNOWLEDGMENTS

We gratefully acknowledge the suggestions from the reviewers to improve this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.539081/full#supplementary-material>

REFERENCES

- Ackers, G. K., Johnson, A. D., and Shea, M. A. (1982). Quantitative model for gene regulation by lambda phage repressor. *Proc. Natl. Acad. Sci. U.S.A.* 79, 1129–33. doi: 10.1073/pnas.79.4.1129
- Bassen, D. M., Vilkhovoy, M., Minot, M., Butcher, J. T., and Varner, J. D. (2017). Jupoets: a constrained multiobjective optimization approach to estimate biochemical model ensembles in the Julia programming language. *BMC Syst. Biol.* 11:10. doi: 10.1186/s12918-016-0380-2
- Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: a fresh approach to numerical computing. *SIAM Rev.* 59, 65–98. doi: 10.1137/141000671
- Bonneau, R., Reiss, D. J., Shannon, P., Facciotti, M., Hood, L., Baliga, N. S., et al. (2006). The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol.* 7:R36. doi: 10.1186/gb-2006-7-5-r36
- Borsook, H. (1950). Protein turnover and incorporation of labeled amino acids into tissue proteins *in vivo* and *in vitro*. *Physiol. Rev.* 30, 206–219. doi: 10.1152/physrev.1950.30.2.206
- Brewster, R. C., Jones, D. L., and Phillips, R. (2012). Tuning promoter strength through rna polymerase binding site design in *Escherichia coli*. *PLoS Comput. Biol.* 8:e1002811. doi: 10.1371/journal.pcbi.1002811
- Chandrasekaran, S., and Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17845–50. doi: 10.1073/pnas.1005139107
- Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J., and Palsson, B. O. (2004). Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 429, 92–6. doi: 10.1038/nature02456
- Covert, M. W., and Palsson, B. O. (2002). Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* 277, 28058–64. doi: 10.1074/jbc.M201691200
- Covert, M. W., Schilling, C. H., and Palsson, B. (2001). Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* 213, 73–88. doi: 10.1006/jtbi.2001.2405
- Doerr, A., de Reus, E., van Nies, P., van der Haar, M., Wei, K., Kattan, J., et al. (2019). Modelling cell-free rna and protein synthesis with minimal systems. *Phys. Biol.* 16:025001. doi: 10.1088/1478-3975/aaf33d
- Flynn, J. M., Neher, S. B., Kim, Y.-I., Sauer, R. T., and Baker, T. A. (2003). Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell* 11, 671–683. doi: 10.1016/S1097-2765(03)00060-1
- Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016). The all *E. coli* TX-TL toolbox 2.0: a platform for cell-free synthetic biology. *ACS Synth. Biol.* 5, 344–55. doi: 10.1021/acssynbio.5b00296
- Garenne, D., Beisel, C. L., and Noireaux, V. (2019). Characterization of the all-*E. coli* transcription-translation system myTXTL by mass spectrometry. *Rapid Commun. Mass Spectrom.* 33, 1036–1048. doi: 10.1002/rcm.8438
- Glass, L., and Kauffman, S. A. (1973). The logical analysis of continuous, non-linear biochemical control networks. *J. Theoret. Biol.* 39, 103–129. doi: 10.1016/0022-5193(73)90208-7
- Gould, R., Bassen, D. M., Chakrabarti, A., Varner, J. D., and Butcher, J. (2016). Population heterogeneity in the epithelial to mesenchymal transition is controlled by NFAT and phosphorylated Sp1. *PLoS Comput. Biol.* 12:e1005251. doi: 10.1371/journal.pcbi.1005251
- Grilly, C., Stricker, J., Pang, W. L., Bennett, M. R., and Hasty, J. (2007). A synthetic gene network for tuning protein degradation in *Saccharomyces cerevisiae*. *Mol. Syst. Biol.* 3, 1–5. doi: 10.1038/msb4100168
- Hoagland, M. B., Keller, E. B., and Zamecnik, P. C. (1956). Enzymatic carboxyl activation of amino acids. *J. Biol. Chem.* 218, 345–358.
- Horvath, N., Vilkhovoy, M., Wayman, J., Calhoun, K., Swartz, J., and Varner, J. (2020). Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in *Escherichia coli*. *Metab. Eng. Commun.* 10:e00113. doi: 10.1016/j.mec.2019.e00113
- Hu, C. Y., Varner, J. D., and Lucks, J. B. (2015). Generating effective models and parameters for RNA genetic circuits. *ACS Synth. Biol.* 4, 914–26. doi: 10.1021/acssynbio.5b00077
- Iizuka, R., Yamagishi-Shirasaki, M., and Funatsu, T. (2011). Kinetic study of *de novo* chromophore maturation of fluorescent proteins. *Anal. Biochem.* 414, 173–8. doi: 10.1016/j.ab.2011.03.036
- Jarontomeechai, T., Stark, J. C., Natarajan, A., Glasscock, C. J., Yates, L. E., Hsu, K. J., et al. (2018). Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nat. Commun.* 9:2686. doi: 10.1038/s41467-018-05620-8
- Karlebach, G., and Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nat. Rev. Mol. Cell Biol.* 9, 770–780. doi: 10.1038/nrm2503
- Karzbrun, E., Shin, J., Bar-Ziv, R. H., and Noireaux, V. (2011). Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett.* 106:048104. doi: 10.1103/PhysRevLett.106.048104
- Kassavetis, G., and Chamberlin, M. (1981). Pausing and termination of transcription within the early region of bacteriophage t7 DNA *in vitro*. *J. Biol. Chem.* 256, 2777–2786.
- Kaufmann, B. B., and van Oudenaarden, A. (2007). Stochastic gene expression: from single molecules to the proteome. *Curr. Opin. Genet. Dev.* 17, 107–12. doi: 10.1016/j.gde.2007.02.007
- Lee, S. B., and Bailey, J. E. (1984). Genetically structured models for lac promoter-operator function in the *Escherichia coli* chromosome and in multicopy plasmids: Lac operator function. *Biotechnol. Bioeng.* 26, 1372–1382. doi: 10.1002/bit.260261115
- Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D., and Arkin, A. P. (2011). Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8617–22. doi: 10.1073/pnas.1015741108
- Maeda, H., Fujita, N., and Ishihama, A. (2000). Competition among seven *Escherichia coli* sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* 28, 3497–503. doi: 10.1093/nar/28.18.3497
- Marshall, R., and Noireaux, V. (2018). “Synthetic biology with an all *E. coli* TXTL system: quantitative characterization of regulatory elements and gene circuits,” in *Synthetic Biology*, ed. J. C. Braman (New York, NY: Humana Press), 61–93. doi: 10.1007/978-1-4939-7795-6_4
- Marshall, R., and Noireaux, V. (2019). Quantitative modeling of transcription and translation of an all-*E. coli* cell-free system. *Sci. Rep.* 9:11980. doi: 10.1038/s41598-019-48468-8
- Matsuura, T., Tanimura, N., Hosoda, K., Yomo, T., and Shimizu, Y. (2017). Reaction dynamics analysis of a reconstituted *Escherichia coli* protein translation system by computational modeling. *Proc. Natl. Acad. Sci. U.S.A.* 114, E1336–E1344. doi: 10.1073/pnas.1615351114
- Matthaei, J. H., and Nirenberg, M. W. (1961). Characteristics and stabilization of DNAase-sensitive protein synthesis in *E. coli* extracts. *Proc. Natl. Acad. Sci. U.S.A.* 47, 1580–8. doi: 10.1073/pnas.47.10.1580
- Mavelli, F., Marangoni, R., and Stano, P. (2015). A simple protein synthesis model for the pure system operation. *Bull. Math. Biol.* 77, 1185–1212. doi: 10.1007/s11538-015-0082-8
- McAdams, H. H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 814–9. doi: 10.1073/pnas.94.3.814
- McClure, W. R. (1980). Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. U.S.A.* 77, 5634–8. doi: 10.1073/pnas.77.10.5634
- Mettetal, J. T., Muzzey, D., Pedraza, J. M., Ozbudak, E. M., and van Oudenaarden, A. (2006). Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7304–9. doi: 10.1073/pnas.0509874103
- Milo, R., Jorgensen, P., Moran, U., Weber, G., and Springer, M. (2010). Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38, D750–D753. doi: 10.1093/nar/gkp889
- Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012). Genetic programs constructed from layered logic gates in single cells. *Nature* 491, 249–53. doi: 10.1038/nature11516
- Moore, S. J., MacDonald, J. T., and Freemont, P. S. (2017). Cell-free synthetic biology for *in vitro* prototype engineering. *Biochem. Soc. Trans.* 45, 785–791. doi: 10.1042/BST20170011
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics.* 33, 161–174. doi: 10.1080/00401706.1991.10484804
- Ng, P. P., Jia, M., Patel, K. G., Brody, J. D., Swartz, J. R., Levy, S. et al. (2012). A vaccine directed to b cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14526–14531. doi: 10.1073/pnas.1211018109

- Niederholtmeyer, H., Sun, Z. Z., Hori, Y., Yeung, E., Verpoorte, A., Murray, R. M., et al. (2015). Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* 4:e09771. doi: 10.7554/eLife.09771
- Nilsson, B. (1984). Probable *in vivo* induction of differentiation by retinoic acid of promyelocytes in acute promyelocytic leukaemia. *Br. J. Haematol.* 57, 365–71. doi: 10.1111/j.1365-2141.1984.tb02910.x
- Nirenberg, M. W., and Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* 47, 1588–1602. doi: 10.1073/pnas.47.10.1588
- Orth, J. D., Fleming, R. M. T., and Palsson, B. Ø. (2010). Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus* 4, 1–47. doi: 10.1128/ecosalplus.10.2.1
- Park, J. O., Rubin, S. A., Xu, Y.-F., Amador-Noguez, D., Fan, J., Shlomi, T., et al. (2016). Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. *Nat. Chem. Biol.* 12, 482–9. doi: 10.1038/nchembio.2077
- Prangemeier, T., Lehr, F.-X., Schoeman, R. M., and Koeppl, H. (2020). Microfluidic platforms for the dynamic characterisation of synthetic circuitry. *Curr. Opin. Biotechnol.* 63, 167–176. doi: 10.1016/j.copbio.2020.02.002
- Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A., and Murali, T. M. (2020). Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat Methods* 17, 147–154. doi: 10.1038/s41592-019-0690-6
- Rackauckas, C., and Nie, Q. (2017). DifferentialEquations.jl—A performant and feature-rich ecosystem for solving differential equations in Julia. *J. Open Res. Soft.* 5:15. doi: 10.5334/jors.151
- Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–26. doi: 10.1016/j.cell.2008.09.050
- Shin, J., and Noireaux, V. (2010). Study of messenger rna inactivation and protein degradation in an *Escherichia coli* cell-free expression system. *J. Biol. Eng.* 4:9. doi: 10.1186/1754-1611-4-9
- Silverman, A. D., Karim, A. S., and Jewett, M. C. (2019). Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* 38, 7364–7377. doi: 10.1093/nar/gkq617
- Sniegowski, J. A., Lappe, J. W., Patel, H. N., Huffman, H. A., and Wachter, R. M. (2005). Base catalysis of chromophore formation in arg96 and glu222 variants of green fluorescent protein. *J. Biol. Chem.* 280, 26248–55. doi: 10.1074/jbc.M412327200
- Soltani, M., Davis, B. R., Ford, H., Nelson, J. A. D., and Bundy, B. C. (2018). Reengineering cell-free protein synthesis as a biosensor: biosensing with transcription, translation, and protein-folding. *Biochem. Eng. J.* 138, 165–171. doi: 10.1016/j.bej.2018.06.014
- Stark, J. C., Jaroentomechai, T., Moeller, T. D., Dubner, R. S., Hsu, K. J., Stevenson, T. C., et al. (2019). On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. *bioRxiv.* 1–41. doi: 10.1101/681841
- Stögbauer, T., Windhager, L., Zimmer, R., and Rädler, J. O. (2012). Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integr. Biol.* 4, 494–501. doi: 10.1039/c2ib00102k
- Swartz, J. R. (2018). Expanding biological applications using cell-free metabolic engineering: an overview. *Metab. Eng.* 50, 156–172. doi: 10.1016/j.ymben.2018.09.011
- Tapia-Rojo, R., Mazo, J. J., Hernández, J. Á., Peleato, M. L., Fillat, M. F., and Faló, F. (2014). Mesoscopic model and free energy landscape for protein-DNA binding sites: analysis of cyanobacterial promoters. *PLoS Comput. Biol.* 10:e1003835. doi: 10.1371/journal.pcbi.1003835
- Tapia-Rojo, R., Prada-Gracia, D., Mazo, J. J., and Faló, F. (2012). Mesoscopic model for free-energy-landscape analysis of DNA sequences. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 86(2 Pt 1):021908. doi: 10.1103/PhysRevE.86.021908
- Tasseff, R., Jensen, H. A., Congleton, J., Dai, D., Rogers, K. V., Sagar, A., et al. (2017). An effective model of the retinoic acid induced hl-60 differentiation program. *Sci. Rep.* 7:14327. doi: 10.1038/s41598-017-14523-5
- Thiery, J. P. (2003). Epithelial-mesenchymal transitions in development and pathologies. *Curr. Opin. Cell Biol.* 15, 740–6. doi: 10.1016/j.ccb.2003.10.006
- Underwood, K. A., Swartz, J. R., and Puglisi, J. D. (2005). Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. *Biotech. Bioeng.* 91, 425–35. doi: 10.1002/bit.20529
- Vilkhovoy, M., Adhikari, A., Vadhin, S., and Varner J. D. (2020). The evolution of cell free biomanufacturing. *Processes* 8:675. doi: 10.3390/pr8060675
- Vilkhovoy, M., Dai, D., Vadhin, S., Adhikari, A., and Varner, J. D. (2019). Absolute quantification of cell-free protein synthesis metabolism by reversed-phase liquid chromatography-mass spectrometry. *J. Vis. Exp.* 152,e60329. doi: 10.3791/60329
- Vilkhovoy, M., Horvath, N., Shih, C.-H., Wayman, J. A., Calhoun, K., Swartz, J., et al. (2018). Sequence specific modeling of *E. coli* cell-free protein synthesis. *ACS Synth. Biol.* 7, 1844–1857. doi: 10.1021/acssynbio.7b00465
- Vilkhovoy, M., Minot, M., and Varner, J. D. (2016). Effective dynamic models of metabolic networks. *IEEE Life Sci. Lett.* 2, 51–54. doi: 10.1109/LLS.2016.2644649
- Winnick, T. (1950a). Studies on the mechanism of protein synthesis in embryonic and tumor tissues. I. evidence relating to the incorporation of labeled amino acids into protein structure in homogenates. *Arch. Biochem.* 27, 65–74.
- Winnick, T. (1950b). Studies on the mechanism of protein synthesis in embryonic and tumor tissues. II. inactivation of fetal rat liver homogenates by dialysis, and reactivation by the adenylic acid system. *Arch. Biochem.* 28, 338–47.
- Yelleswarapu, M., Linden, A. J. V. D., Sluijs, B. V., Pieters, P. A., Dubuc, E., Greef, T. F. A. D., et al. (2018). Sigma factor-mediated tuning of bacterial cell-free synthetic genetic oscillators. *ACS Synth. Biol.* 7, 2879–2887. doi: 10.1021/acssynbio.8b00300
- Yue, K., Zhu, Y., and Kai, L. (2019). Cell-free protein synthesis: chassis toward the minimal cell. *Cells* 8:315. doi: 10.3390/cells8040315

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Adhikari, Vilkhovoy, Vadhin, Lim and Varner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.