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 14853

Running Title: Effective Transcription and Translation Models

Submitted: *Frontiers in Bioengineering and Biotechnology*

*Corresponding author:

Jeffrey D. Varner,

Professor, School of Chemical and Biomolecular Engineering,

244 Olin Hall, Cornell University, Ithaca NY, 14853

Email: jdv27@cornell.edu

Phone: (607) 255 - 4258

Fax: (607) 255 - 9166

Abstract

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. We tested this approach by simulating the dynamics of two cell free synthetic circuits. 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. While many of the model parameters were estimated from previous biophysical literature, the remaining unknown model parameters for each circuit were estimated from messenger RNA (mRNA) and protein measurements using multiobjective optimization. In particular, either the literature parameter estimates were used directly in the model simulations, or characteristic literature values were used to establish feasible ranges for the multiobjective parameter search. Next, global sensitivity analysis was used to determine the influence of individual model parameters on the expression dynamics. 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. The model code, parameters, and analysis scripts are available for download under an MIT software license from the Varnerlab GitHub repository.

Introduction

Transcription (TX) and translation (TL), 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 heath. For example, evolutionarily conserved developmental programs such as the epithelial to mesenchymal transition (EMT) [1], or retinoic acid induced differentiation [2], 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 regulatory networks which control the metabolic state of the cell in response to changing nutrient conditions [3, 4]. 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 [5]. Logical models such as Boolean networks [6] developed using a variety of approaches and data [7] 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 [8]. Lastly, stochastic models describe the interactions between individual molecules, and discrete reaction events [9–12]. 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., [13], 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 [14–16], and the role of adenosine triphosphate (ATP) in protein production [17]. Further, *E. coli* extracts were used by Nirenberg and Matthaei in 1961 to demonstrate templated translation [18, 19], leading to a Nobel Prize in 1968 for deciphering the codon code. More recently, as advancements in extract preparation and energy regeneration have extend their durability, the usage of cell free systems has also expanded to both small- and large-scale biotechnology and biomanufacturing applications [20, 21]. Today, cell free systems have been implemented for therapeutic protein and vaccine production [22–24], biosensing [25], genetic part prototyping [26] and minimal cell systems [27]. 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 [28–32]. However, despite the obvious advantages offered by a cell free system, experimental determination of the kinetic parameters for these models is often not feasible. 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 [33]. 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. We tested this approach by simulating the

dynamics of two cell free synthetic circuits. First, we considered a simple circuit ($\mathcal{C}1$) 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 ($\mathcal{C}2$) where two control genes were coupled to the expression of deGFP. Characteristic values for many of the model parameters were estimated from previous biophysical literature, while the remaining unknown model parameters for each circuit were estimated from messenger RNA (mRNA) and protein measurements using multiobjective optimization. In particular, either the literature parameter estimates were used directly in the model simulations, or characteristic literature values were used to establish feasible ranges for the multiobjective parameter search. Next, Morris sensitivity analysis was used to determine the influence of individual model parameters on the expression dynamics of each circuit. For $\mathcal{C}1$, the sensitivity results were informative, but expected. However, for $\mathcal{C}2$, the analysis hierarchically stratified the parameters (and associated model species) into local versus global categories. For example, parameters that controlled the abundance of lambda phage repressor protein (cI-ssrA), a master circuit regulator in $\mathcal{C}2$, were globally important as they influenced all other species. On the other hand, the parameters that influenced deGFP levels (the endpoint of both circuits) were only locally important to deGFP. Taken together, the effective biophysical modeling approach captured the expression dynamics, including the transcription dynamics, for two synthetic cell free circuits. While we considered only two circuits here, this approach could potentially be extended to simulate other genetic circuits in both cell free and whole cell biomolecular applications. The model code, parameters, and analysis scripts are available under an MIT software license from the Varnerlab GitHub repository [34].

Materials and Methods

myTX/TL cell-free protein synthesis. The cell-free protein synthesis (CFPS) reactions were carried out using the myTXTL Sigma 70 Master Mix (Arbor Biosciences) in 1.5 mL Eppendorf tubes. The working volume of all the reactions was 12 μL, composed of the Sigma 70 Master Mix (9 μL) and the plasmids (3 μL total): P70a-deGFP (5 nM) for the single-gene system; P70a-deGFP-ssrA (8 nM), P70a-S28 (1.5 nM) and P28a-cI-ssrA (1 nM) for the negative feedback circuit. The plasmids were bought in lyophilized form (Arbor Biosciences) and purified using QIAprep Spin Miniprep Kit (Qiagen) using cell lines DH5-Alpha (for P28a-cI-ssrA) or KL740 (for P70a-deGFP, P70a-deGFP-ssrA, and P70a-S28). The CFPS reactions were incubated at 29°C.

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

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

Formulation and solution of model equations Consider a gene regulatory motif with genes $\mathcal{G} = 1, 2, \ldots, \mathcal{N}$. Each node in this motif is governed by two differential equations, one for mRNA (m_j) and the corresponding protein (p_j):

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

$$\dot{p}_j = r_{L,j} w_j - \theta_{p,j} p_j \tag{2}$$

The term $r_{X,j} u_j$ in the mRNA balance, which denotes the rate of transcription for gene j, is the product of a kinetic limit $r_{X,j}$ (nmol h^{-1}) and a transcription control function $0 \leq u_j(\ldots) \leq 1$ (dimensionless). Similarly, the rate of translation of mRNA j, denoted by $r_{L,j} w_j$, is also the product of the kinetic limit of translation and a translational control term $0 \leq w_j(\ldots) \leq 1$ (dimensionless). Lastly, $\theta_{\star,j}$ denotes the first-order rate constant (h^{-1}) governing degradation of protein and mRNA. The model equations, encoded in the Julia programming language [35], were automatically generated using the JuGRN tool [36]. The model equations were solved numerically using the `Rosenbrock23` routine of the `DifferentialEquations.jl` package [37].

Derivation of the transcription and translation kinetic limits. The key idea behind the derivation of the kinetic limit of transcription and translation is that the polymerase (or ribosome) acts as a pseudo-enzyme; it binds a substrate (gene or message), reads the gene (or message), and then dissociates. Thus, we might expect that we could use a strategy similar to enzyme kinetics to derive an expression for $r_{X,j}$ (or $r_{L,j}$). Toward this idea, the kinetic limit of transcription $r_{X,j}$ was developed by proposing a simple four

elementary step model:

$$
\begin{aligned}
\mathcal{G}_j + R_X^{\circ} &\rightleftharpoons (\mathcal{G}_j : \mathrm{R_X})_C \\
(\mathcal{G}_j : \mathrm{R_X})_C &\longrightarrow (\mathcal{G}_j : \mathrm{R_X})_O \\
(\mathcal{G}_j : \mathrm{R_X})_O &\longrightarrow R_X + \mathcal{G}_j \\
(\mathcal{G}_j : \mathrm{R_X})_O &\longrightarrow m_j + R_X + \mathcal{G}_j
\end{aligned}
$$

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

$$r_{X,j} = k_{E,j} \left(\mathcal{G}_j : \mathrm{R_X}\right)_O \tag{3}$$

where $k_{E,j}$ is the elongation rate constant for gene j. Then the functional form for $r_{X,j}$ can be derived:

$$r_{X,j} = V_{X,j}^{max} \left(\frac{\mathcal{G}_j}{\tau_{X,j} K_{X,j} + \left(1 + \tau_{X,j}\right) \mathcal{G}_j + \mathcal{O}_{X,j}} \right) \tag{4}$$

where $V_{X,j}^{max}$ denotes the maximum transcription rate (nM/h) of gene j and:

$$\mathcal{O}_{X,j} = \sum_{i=1,j}^{\mathcal{N}} \frac{K_{X,j}\tau_{X,j}}{K_{X,i}\tau_{X,i}} \left(1 + \tau_{X,i}\right) \mathcal{G}_i \tag{5}$$

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 a model for the translational kinetic limit:

$$r_{L,j} = V_{L,j}^{max} \left(\frac{m_j}{\tau_{L,j} K_{L,j} + \left(1 + \tau_{L,j}\right) m_j + \mathcal{O}_{L,j}} \right) \tag{6}$$

where $V_{L,j}^{max}$ denotes the maximum translation rate (nM/h) and:

$$\mathcal{O}_{L,j} = \sum_{i=1,j}^{\mathcal{N}} \frac{K_{L,j}\tau_{L,j}}{K_{L,i}\tau_{L,i}} \left(1 + \tau_{L,i}\right) m_i \tag{7}$$

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 were defined as $K_{\star,j}^{-1} \equiv k_{+,j}/\left(k_{-,j} + k_{I,j}\right)$ and $\tau_{\star,j}^{-1} \equiv k_{I,j}/\left(k_{A,j} + k_{E,j}\right)$, respectively. 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:

$$V_{X,j}^{max} \equiv \left[R_{X,T}\left(\frac{\dot{v}_X}{l_{G,j}}\right)\right] \tag{8}$$

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

$$V_{L,j}^{max} \equiv \left[K_P R_{L,T}\left(\frac{\dot{v}_L}{l_{P,j}}\right)\right] \tag{9}$$

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

Control functions u and w. Values of the control functions $u\left(\ldots\right)$ and $w\left(\ldots\right)$ describe the regulation of transcription and translation. Ackers et al., borrowed from statistical mechanics to recast the $u\left(\ldots\right)$ function as the probability that a system exists in a configuration which leads to expression [38]. The idea of recasting $u\left(\ldots\right)$ as the probability of expression was also developed (apparently independently) by Bailey and coworkers in a series

of papers modeling the *lac* operon, see [39]. More recently, Moon and Voigt adapted a similar approach when modeling the expression of synthetic circuits in *E. coli* [40]. The $u(\ldots)$ function is formulated as:

$$u(\ldots)_j = \frac{\sum_{i \in \{\chi\}} W_i f_i(\ldots)}{\sum_{j \in \mathcal{C}_j} W_j f_j(\ldots)} \tag{10}$$

where W_i (dimensionless) denotes the weight of configuration i, while $f_i(\cdots)$ (dimensionless) is a binding function (taken to be a hill-type function) describing the fraction of bound activator/inhibitor for configuration i. The summation in the numerator of Eq. (10) is over the set of configurations that lead to expression (denoted as χ), while the summation in the numerator is over the set of all possible configurations for gene j (denoted as $\mathcal{C}_j$). Thus, $u(\ldots)_j$ can be thought of as the fraction of all possible configurations that lead to expression. The weights W_i are related to the Gibbs energy of configuration i: W_i = $\exp(-\Delta G_i/RT)$ where ΔG_i denotes the molar Gibbs energy for configuration i (kJ/mol), R denotes the ideal gas constant (kJ mol^{-1} K^{-1}), and T denotes the system temperature (Kelvin) [38].

The translational control function $w(\ldots)$ described the loss of cell free translational activity. Loss of translational activity could be a function of many factors, including depletion of metabolic resources. However, in this study we assumed a translational activity factor ϵ of the form:

$$\dot{\epsilon} = -\left(\frac{0.693}{\tau_{L,1/2}}\right)\epsilon \tag{11}$$

where $\tau_{L,1/2}$ denotes the half-life of translation (estimated from the experimental data) and $\epsilon(0)$ = 1. The translational control variable was then given by $w_i = \epsilon$ for all translation processes.

Estimation of model parameters Model parameters were estimated using the Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) [41]. Several of the model parameters were estimated directly from literature, or were set by experimental conditions (Table 1). However, the remaining unknown parameters ($\mathbf{k}$) were estimated from mRNA and protein concentration measurements using JuPOETs. JuPOETs is a multiobjective optimization approach which integrates simulated annealing with Pareto optimality to estimate parameter ensembles on or near the optimal tradeoff surface between competing training objectives. JuPOETs minimized training objectives of the form:

$$\mathcal{E}_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j} \left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k}) \right)^2 \tag{12}$$

subject to the model equations, initial conditions, parameter bounds $\mathcal{L} \leq \mathbf{k} \leq \mathcal{U}$ and maximum value constraints for some unmeasured protein species. The objective function(s) $\mathcal{E}_j$ measured the squared difference between the model simulations and experimental measurements for training objective j. The symbol $\hat{\mathcal{M}}_{ij}$ denotes an experimental observation at time index i for training objective j, while the symbol $\hat{y}_{ij}$ denotes the model simulation output at time index i from training objective 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 ($\mathcal{C}1$), $\mathcal{E}_1$ corresponded to mRNA deGFP, while $\mathcal{E}_2$ corresponded to the deGFP protein concentration. On the other hand, for the negative feedback model ($\mathcal{C}2$), $\mathcal{E}_1$ 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 non-physical accumulation of the cI-ssrA protein (unmeasured) with a term of the form: $\mathcal{E}_5 = C \times \max(0, \hat{y} - U)$ where C denotes a penalty parameter (C = 1$\times 10^5$), $\hat{y}$ denotes the maximum cI-ssrA concentration, and U denotes a concentration upper bound (U = 100nM). For $\mathcal{C}1$, 11 parameters were estimated, while 33 parameters were estimated for $\mathcal{C}2$.

Parameter bounds $\mathcal{L} \leq \mathbf{k} \leq \mathcal{U}$ were established from literature, or from previous experimental or 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.

Morris sensitivity analysis Morris sensitivity analysis was used to understand which model parameters were sensitive [42]. 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, over a range of values for each parameter; the mean of elementary effects measures the direct effect of a particular parameter on the performance function, while the variance of each elementary effect indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests connectedness or nonlinearity). The Morris sensitivity coefficients were computed using the `DiffEqSensitivity.jl` package [37]. 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. The Morris sensitivity coefficients are calculated using the `compute_sensitivity_coefficients.jl` script in the model repositories.

Results

Synthetic circuit architecture The two genetic circuits ($\mathcal{C}1$ and $\mathcal{C}2$) used in this study were based upon the bacterial sigma factor regulatory system (Fig. 1). Sigma factor 70 (σ_{70}), endogenously present in the extract, was the primary driver of each circuit. In $\mathcal{C}1$, σ_{70} induced green fluorescent protein (deGFP) expression was explored in the absence of additional regulators or protein degradation (Fig. 1A). In $\mathcal{C}2$, σ_{70} induced the expression of sigma factor 28 (σ_{28}) and deGFP-ssrA (Fig. 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 [43]. 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 [44, 45]. In addition, messenger RNAs (mRNAs) were always subject to degradation due to the presence of degradation enzymes in the extract [45, 46]. Taken together, the interactions of the components manifested in an accumulation of deGFP protein for $\mathcal{C}1$, and a pulse signal of deGFP-ssrA in $\mathcal{C}2$. Studying $\mathcal{C}1$ allowed us to estimate parameters governing the interaction of σ_{70} with the P70a promoter. Whereas, the $\mathcal{C}2$ 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.

Modeling and analysis of the $\mathcal{C}1$ 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 $\mathcal{C}1$ (Fig. 2). JuPOETs produced an ensemble (N = 140) of the 11 unknown model parameters which captured the transcription of mRNA (Fig. 2A) and the translation of deGFP protein (Fig. 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 two hours, and stayed at this level for the remainder of the reaction. Thus, the cell free reaction maintained continuous transcriptional activity with an average mRNA lifetime of 27 minutes; Garamella et al [45] reported a similar lifetime of 17 - 18 min. On the other hand, deGFP protein concentration increased more slowly, and began to saturate between 8 to 10 hr at approximately 15 μM. Given there was negligible protein degradation (the mean deGFP half-life was estimated to be $\sim$11 days, which was similar to the value of 6 days estimated by Horvath et al, albeit in a different cell free system [33]). The saturating protein concentration suggested that the translational capacity of the cell free system decreased over the course of the reaction. The decrease in translational capacity, which could stem from several sources, was captured in the simulations using a monotonically decreasing translation capacity state variable ϵ, and the translational control variable $w\left(\ldots\right)$. In particular, the mean half-life of translational capacity was estimated to be $\tau_{L,1/2} \sim$ 4 h in the $\mathcal{C}1$ experiments. Taken together, JuPOETs produced an ensemble of model parameters that captured the experimental training data. Next, we considered which $\mathcal{C}1$ model parameters were important to the model performance using Morris sensitivity analysis, a global sensitivity analysis method.

The importance of $\mathcal{C}1$ model parameters was quantified using Morris sensitivity analysis (Fig. 2B). The Morris method computes the influence of each parameter, known as the elementary effect, on a model performance function. The mean of elementary effects measures the direct effect of a particular parameter, while the variance indicates whether the effects are non-linear or the result of interactions with other parameters (large variance suggests nonlinearity). The performance function was defined as the area under the curve for each species. 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, transla-

tional 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 $\mathcal{C}1$ model parameters highlighted the hierarchical structure of the transcriptional and translation 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 $\mathcal{C}1$ to constrain the analysis of $\mathcal{C}2$.

Modeling and analysis of the $\mathcal{C}2$ circuit The effective biophysical transcription and translation model captured the deGFP-ssrA expression dynamics in the negative feedback circuit $\mathcal{C}2$ (Fig. 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 $\mathcal{C}1$, the $\mathcal{C}2$ 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 $\mathcal{C}1$, 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 $\mathcal{C}1$, the regulated mRNA concentrations reached an approximate steady-state. This further confirmed continuous transcription and mRNA degradation throughout the cell free reaction. The mean estimated mRNA lifetime for cI-ssrA and deGFP were similar (approximately 16 min), while the degradation of σ_{28} mRNA was predicted to be slower (mean mRNA lifetime was estimated to be approximately 30 min). Lastly, the mean peak degradation rate for GFP was approximately 47 nM/min, while the mean peak cI-ssrA degradation rate was predicted to be approximately 63 nM/min; both of these degradation rate estimates were consistent with the previously reported range of 15 nM/min - 150 nM/min measured by Garamella et al [45].

The secondary effect of cI-ssrA repression was visible in the cI-ssrA mRNA expression pattern. The expression of cI-ssrA was induced by σ_{28}, however, σ_{28} expression was repressed by cI-ssrA, thereby completing a negative feedback loop. Initially, before appreciable levels of cI-ssrA had been translated, the cI-ssrA transcription rate was maximum (approximately 200 nM/h). However, the transcription rate decreased to approximately 12 nM/h after two hours and remained constant for the remainder of the cell free reaction. Similarly, transcription rates for σ_{28} (approximately 1200 nM/h) and deGFP-ssrA (approximately 750 nM/h) were initially at a maximum due to the presence of endogenous σ_{70}, but then quickly dropped as cI protein accumulated. Protein synthesis followed a similar trend, with the translation rates for σ_{28} and deGFP-ssrA initially present at their maximum values before quickly dropping. After one hour, deGFP levels reached a peak and decayed due to the ClpXP-mediated degradation, whereas σ_{28} protein levels continued to slowly rise at a steady rate (approximately 15 nM/h). The $\mathcal{C}2$ 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 $\mathcal{C}2$. Next, we considered which $\mathcal{C}2$ model parameters were important using Morris sensitivity analysis.

Morris sensitivity analysis of the negative feedback circuit $\mathcal{C}2$ stratified the parameters into locally and globally important groups (Fig. 3B). The influence of 33 parameters was computed using the area under the curve 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 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 $\mathcal{C}2$ 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.

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 ($\mathcal{C}1$ and $\mathcal{C}2$). 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 [47]. The RNAP and ribosome elongation rates were taken from Garamella et al. [45], while other parameter were estimated from BioNumbers [48]. Similarly, the weights appearing in the transcription control function $u(...)$ were based upon the Gibbs energies of the respective promoter configurations, while the form of the transcriptional control functions was derived from a statistical mechanical treatment of promoter activity [38–40]. 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 $\mathcal{C}1$, 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 $\mathcal{C}2$, which involved deGFP expression subject to negative feedback and programmed protein degradation. We estimated which model parameters were important to the performance of $\mathcal{C}1$ and $\mathcal{C}2$ using Morris sensitivity analysis. Sensitivity analysis results for $\mathcal{C}1$ 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 $\mathcal{C}2$ were more nuanced, with parameters (and associated species) being stratified into locally and globally important groups; the performance of $\mathcal{C}2$ 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 [49]. Boolean and probabilisitic approaches [50–52] 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 course 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 interesting application could be a description of the integration of gene expression with metabolism in cell free synthetic biology applications. For example, 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. We have recently started to explore this question by integrating effective transcription and translation models with metabolic networks in cell free reactions e.g., [33, 53], and also developing experimental tools to measure metabolite concentrations in cell free systems [54]. However, these previous transcriptional and translations models (and similar precursor models simulating eukaryotic processes [55, 56]) 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.

The effective TX/TL model described the experimental mRNA and protein training data. However, there were several important questions to be addressed by future studies. First, the model formulation described the data, but did not predict dynamics outside of

the training set. If this approach is to be useful to the synthetic biology community, or more broadly as an effective biophysical technique to model *in vivo* gene expression dynamics for applications such as regulatory flux balance analysis, we need to have confidence that the modeling approach is predictive. Thus, while we have established a descriptive model, we have to yet to establish a predictive model. Next, there were several technical or mechanistic questions that should be explored further. For example, cI-ssrA represses the activity of the P70a promoter via interaction with its OR2 and OR1 operator sites; in this study we considered only a single operator site suggesting that we potentially underestimated the potency of cI repression in the deGFP and σ_{28} promoter functions, see the multiplication rule [57]. Further, we used a first order approximation of ClpXP mediated protein degradation, while Garamella et al. [45] 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 [32]. For example, one of the potential limitations of the current study (that was addressed by Marshall and Noireaux [32]) is that we did not consider a separate species for dark GFP. In our previous RNA circuit modeling [58], we did include this term, but failed to do so here. We expect inclusion of a dark versus light GFP species could influence the values for the estimated parameters, particularly the translation time constants. However, previous reports suggested the *in vitro* maturation time of deGFP was approximately 8 min [59], much faster than the typical maturation times for GFP of one hour *in vivo* [60, 61]. Thus, the impact of including a dark versus light GFP species may not be worth the increased model complexity. Lastly, we should validate the values estimated for the binding function parameters and the promoter configuration free energies. Maeda et al measured the binding affinities of the seven *E. coli* σ factors with RNAP [62]; 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., [38, 63–65]. 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.

Conflict of Interest Statement

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.

Author Contributions

J.V directed the study. M.V, S.V, H.L and A.A conducted the cell free experimental measurements. J.V, M.V and A.A developed the reduced order models and the parameter ensemble. M.V, A.A and J.V analyzed the model ensemble, and generated figures for the manuscript. The manuscript was prepared and edited for publication by A.A, M.V, S.V and J.V. 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 anonymous reviewers to improve this manuscript.

Data Availability Statement

Model code is available under an MIT software license from the Varnerlab GitHub repository [34]. 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.

References

1. Thiery JP (2003) Epithelial-mesenchymal transitions in development and pathologies. Curr Opin Cell Biol 15: 740-6.
2. Nilsson B (1984) Probable in vivo induction of differentiation by retinoic acid of promyelocytes in acute promyelocytic leukaemia. Br J Haematol 57: 365-71.
3. Orth JD, Fleming RMT, Palsson BØ (2010) Reconstruction and use of microbial metabolic networks: the core escherichia coli metabolic model as an educational guide. EcoSal Plus 4.
4. Vilkhovoy M, Minot M, Varner JD (2016) Effective dynamic models of metabolic networks. IEEE Life Sciences Letters 2: 51-54.
5. Karlebach G, Shamir R (2008) Modelling and analysis of gene regulatory networks. Nature Reviews Molecular Cell Biology 9: 770–780.
6. Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. Journal of theoretical Biology 39: 103–129.
7. Pratapa A, Jalihal AP, Law JN, Bharadwaj A, Murali TM (2020) Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. Nat Methods 17: 147-154.
8. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, et al. (2006) The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 7: R36.
9. McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A 94: 814-9.
10. Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. Proc Natl Acad Sci U S A 103: 7304-9.
11. Kaufmann BB, van Oudenaarden A (2007) Stochastic gene expression: from single

molecules to the proteome. Curr Opin Genet Dev 17: 107-12.

12. Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135: 216-26.
13. Park JO, Rubin SA, Xu YF, Amador-Noguez D, Fan J, et al. (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat Chem Biol 12: 482-9.
14. Borsook H (1950) Protein turnover and incorporation of labeled amino acids into tissue proteins in vivo and in vitro. Physiological reviews 30: 206–219.
15. Winnick T (1950) 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.
16. Winnick T (1950) 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.
17. Hoagland MB, Keller EB, Zamecnik PC (1956) Enzymatic carboxyl activation of amino acids. J Biol Chem 218: 345–358.
18. Nirenberg MW, Matthaei JH (1961) The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci 47: 1588-602.
19. Matthaei JH, Nirenberg MW (1961) Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. Proc Natl Acad Sci 47: 1580-8.
20. Swartz JR (2018) Expanding biological applications using cell-free metabolic engineering: An overview. Metabolic engineering 50: 156–172.
21. Silverman AD, Karim AS, Jewett MC (2019) Cell-free gene expression: an expanded repertoire of applications. Nature Reviews Genetics : 1–20.
22. Jaroentomeechai T, Stark JC, Natarajan A, Glasscock CJ, Yates LE, et al. (2018)

Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. Nature communications 9: 2686.

23. Stark JC, Jaroentomeechai T, Moeller TD, Dubner RS, Hsu KJ, et al. (2019) On-demand, cell-free biomanufacturing of conjugate vaccines at the point-of-care. bioRxiv : 681841.
24. Ng PP, Jia M, Patel KG, Brody JD, Swartz JR, et al. (2012) A vaccine directed to b cells and produced by cell-free protein synthesis generates potent antilymphoma immunity. Proc Natl Acad Sci 109: 14526–14531.
25. Soltani M, Davis BR, Ford H, Nelson JAD, Bundy BC (2018) Reengineering cell-free protein synthesis as a biosensor: Biosensing with transcription, translation, and protein-folding. Biochemical Engineering Journal 138: 165–171.
26. Moore SJ, MacDonald JT, Freemont PS (2017) Cell-free synthetic biology for in vitro prototype engineering. Biochemical Society Transactions 45: 785–791.
27. Yue K, Zhu Y, Kai L (2019) Cell-free protein synthesis: chassis toward the minimal cell. Cells 8: 315.
28. Stögbauer T, Windhager L, Zimmer R, Rädler JO (2012) Experiment and mathematical modeling of gene expression dynamics in a cell-free system. Integrative Biology 4: 494–501.
29. Mavelli F, Marangoni R, Stano P (2015) A simple protein synthesis model for the pure system operation. Bulletin of mathematical biology 77: 1185–1212.
30. Matsuura T, Tanimura N, Hosoda K, Yomo T, Shimizu Y (2017) Reaction dynamics analysis of a reconstituted escherichia coli protein translation system by computational modeling. Proceedings of the National Academy of Sciences 114: E1336–E1344.
31. Doerr A, de Reus E, van Nies P, van der Haar M, Wei K, et al. (2019) Modelling cell-free rna and protein synthesis with minimal systems. Physical biology 16: 025001.

32. Marshall R, Noireaux V (2019) Quantitative modeling of transcription and translation of an all-e. coli cell-free system. Sci Rep 9: 11980.
33. Horvath N, Vilkhovoy M, Wayman J, Calhoun K, Swartz J, et al. (2020) Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in Escherichia coli. Metab Eng Comm 10: e00113.
34. Varnerlab. Github reposistory for tx/tl model code. URL https://github.com/varnerlab/Biophysical-TXTL-Model-Code.
35. Bezanson J, Edelman A, Karpinski S, Shah VB (2017) Julia: A fresh approach to numerical computing. SIAM Review 59: 65–98.
36. Varnerlab. Gene Regulatory Network Generator in Julia (JuGRN). URL https://github.com/varnerlab/JuGRN-Generator.
37. Rackauckas C, Nie Q (2017) DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. J Open Res Soft 5: 15.
38. Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A 79: 1129-33.
39. Lee SB, Bailey JE (1984) Genetically structured models forlac promoter–operator function in the *Escherichia coli* chromosome and in multicopy plasmids: Lac operator function. Biotechnology and Bioengineering 26: 1372–1382.
40. Moon TS, Lou C, Tamsir A, Stanton BC, Voigt CA (2012) Genetic programs constructed from layered logic gates in single cells. Nature 491: 249-53.
41. Bassen DM, Vilkhovoy M, Minot M, Butcher JT, Varner JD (2017) Jupoets: a constrained multiobjective optimization approach to estimate biochemical model ensembles in the julia programming language. BMC Syst Biol 11: 10.
42. Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics : 161–174.
43. Marshall R, Noireaux V (2018) Synthetic biology with an all e. coli txtl system: Quanti-

tative characterization of regulatory elements and gene circuits. In: Synthetic Biology, Springer. pp. 61–93.

44. Flynn JM, Neher SB, Kim YI, Sauer RT, Baker TA (2003) Proteomic discovery of cellular substrates of the clpxp protease reveals five classes of clpx-recognition signals. Molecular cell 11: 671–683.
45. Garamella J, Marshall R, Rustad M, 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.
46. Karzbrun E, Shin J, Bar-Ziv RH, Noireaux V (2011) Coarse-grained dynamics of protein synthesis in a cell-free system. Physical review letters 106: 048104.
47. McClure WR (1980) Rate-limiting steps in rna chain initiation. Proc Natl Acad Sci U S A 77: 5634-8.
48. Milo R, Jorgensen P, Moran U, Weber G, Springer M (2010) Bionumbers–the database of key numbers in molecular and cell biology. Nucleic Acids Res 38: D750-3.
49. Covert MW, Palsson BØ (2002) Transcriptional regulation in constraints-based metabolic models of escherichia coli. J Biol Chem 277: 28058-64.
50. Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213: 73-88.
51. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429: 92-6.
52. Chandrasekaran S, Price ND (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.
53. Vilkhovoy M, Horvath N, Shih CH, Wayman JA, Calhoun K, et al. (2018) Sequence Specific Modeling of *E. coli* Cell-Free Protein Synthesis. ACS Synth Biol 7: 1844-1857.

54. Vilkhovoy M, Dai D, Vadhin S, Adhikari A, Varner JD (2019) Absolute quantification of cell-free protein synthesis metabolism by reversed-phase liquid chromatography-mass spectrometry. J Vis Exp .
55. Gould R, Bassen DM, Chakrabarti A, Varner JD, Butcher J (2016) Population heterogeneity in the epithelial to mesenchymal transition is controlled by nfat and phosphorylated sp1. PLoS Comput Biol 12: e1005251.
56. Tasseff R, Jensen HA, Congleton J, Dai D, Rogers KV, et al. (2017) An effective model of the retinoic acid induced hl-60 differentiation program. Sci Rep 7: 14327.
57. Lucks JB, Qi L, Mutalik VK, Wang D, Arkin AP (2011) Versatile rna-sensing transcriptional regulators for engineering genetic networks. Proc Natl Acad Sci U S A 108: 8617-22.
58. Hu CY, Varner JD, Lucks JB (2015) Generating effective models and parameters for rna genetic circuits. ACS Synth Biol 4: 914-26.
59. Shin J, Noireaux V (2010) Study of messenger rna inactivation and protein degradation in an escherichia coli cell-free expression system. J Biol Eng 4: 9.
60. Sniegowski JA, Lappe JW, Patel HN, Huffman HA, Wachter RM (2005) Base catalysis of chromophore formation in arg96 and glu222 variants of green fluorescent protein. J Biol Chem 280: 26248-55.
61. Iizuka R, Yamagishi-Shirasaki M, Funatsu T (2011) Kinetic study of de novo chromophore maturation of fluorescent proteins. Anal Biochem 414: 173-8.
62. Maeda H, Fujita N, Ishihama A (2000) Competition among seven escherichia coli sigma subunits: relative binding affinities to the core rna polymerase. Nucleic Acids Res 28: 3497-503.
63. Brewster RC, Jones DL, Phillips R (2012) Tuning promoter strength through rna polymerase binding site design in escherichia coli. PLoS Comput Biol 8: e1002811.
64. Tapia-Rojo R, Prada-Gracia D, Mazo JJ, Falo F (2012) Mesoscopic model for free-

energy-landscape analysis of dna sequences. Phys Rev E Stat Nonlin Soft Matter Phys 86: 021908.

65. Tapia-Rojo R, Mazo JJ, Hernández JÁ, Peleato ML, Fillat MF, et al. (2014) Mesoscopic model and free energy landscape for protein-dna binding sites: analysis of cyanobacterial promoters. PLoS Comput Biol 10: e1003835.
66. Underwood KA, Swartz JR, Puglisi JD (2005) Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. Biotech Bioeng 91: 425-35.
67. Kassavetis G, Chamberlin M (1981) Pausing and termination of transcription within the early region of bacteriophage t7 dna in vitro. Journal of Biological Chemistry 256: 2777–2786.
68. Niederholtmeyer H, Sun ZZ, Hori Y, Yeung E, Verpoorte A, et al. (2015) Rapid cell-free forward engineering of novel genetic ring oscillators. Elife 4: e09771.
69. Grilly C, Stricker J, Pang WL, Bennett MR, Hasty J (2007) A synthetic gene network for tuning protein degradation in saccharomyces cerevisiae. Molecular systems biology 3.

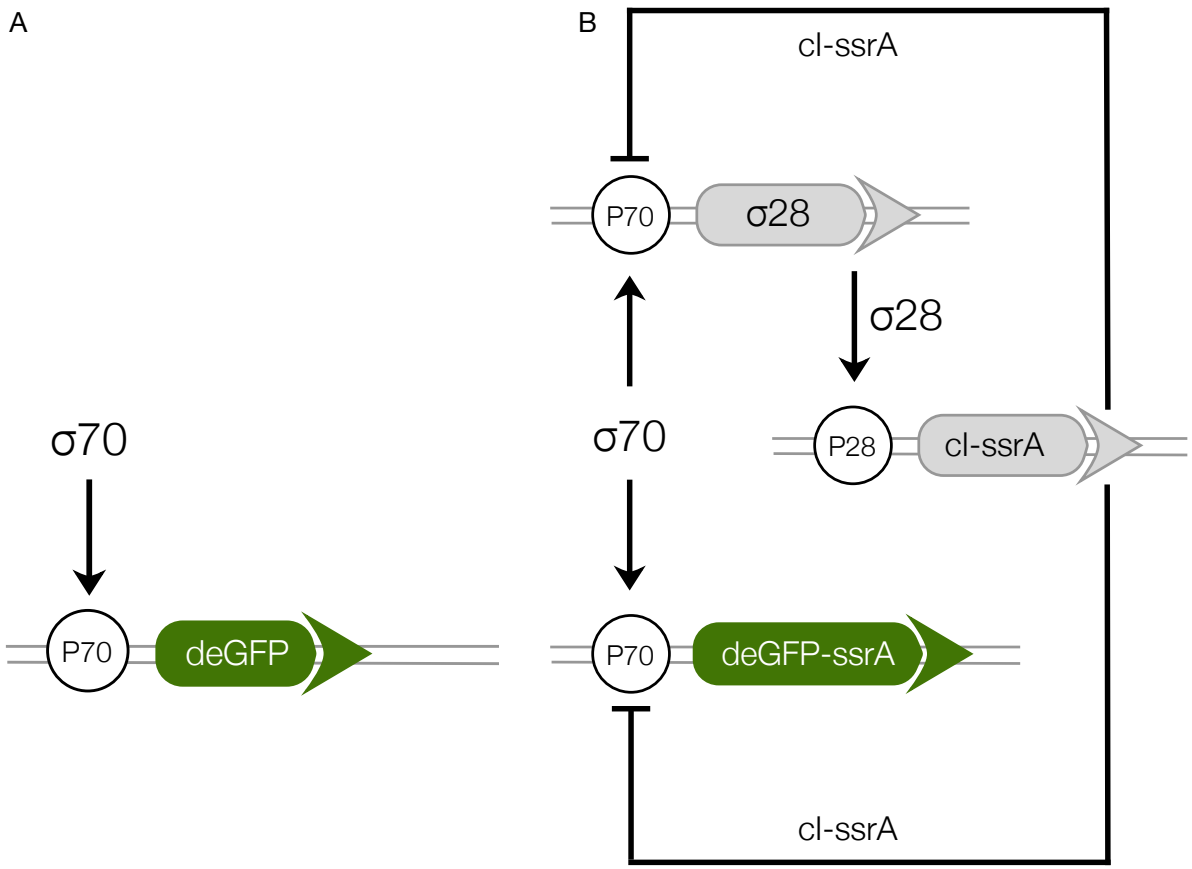

Fig. 1: Schematic of the cell free gene expression circuits used in this study. **A**: Sigma factor 70 (σ_{70}) induced expression of deGFP. **B**: The circuit components encode for a negative feedback loop motif. Sigma factor 28 and deGFP-ssrA genes on the P70a promoters are expressed first because of the endogenous presence of sigma 70 factor in the extract. Sigma factor 28, once expressed, induces the P28a promoter, turning on the expression of the cI-ssrA gene which represses the P70a promoter. The circuit is modified from a previous study [45] by including an ssrA degradation tag on the cI gene.

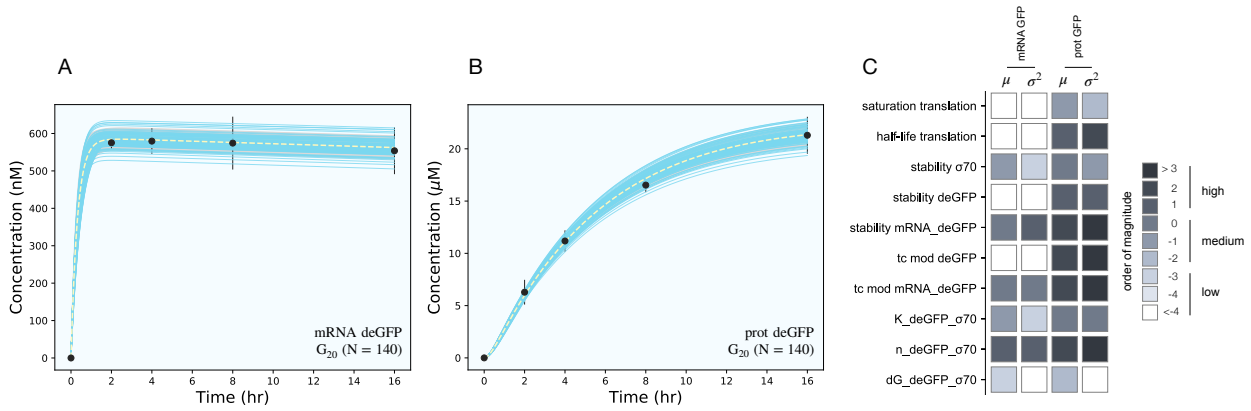

Fig. 2: Model simulations versus experimental measurements for σ_{70} induced deGFP expression. **A**: Simulated and measured deGFP mRNA concentration versus time using the small circuit G_{20} ensemble (N = 140). **B**: Simulated and measured deGFP protein concentration versus time the small circuit G_{20} ensemble (N = 140). **C**: Global sensitivity analysis of the P70-deGFP circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble. Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 140 parameter sets that were rank two or below. The parameter ensemble was used to calculate the mean (dashed line) and the 95% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates.

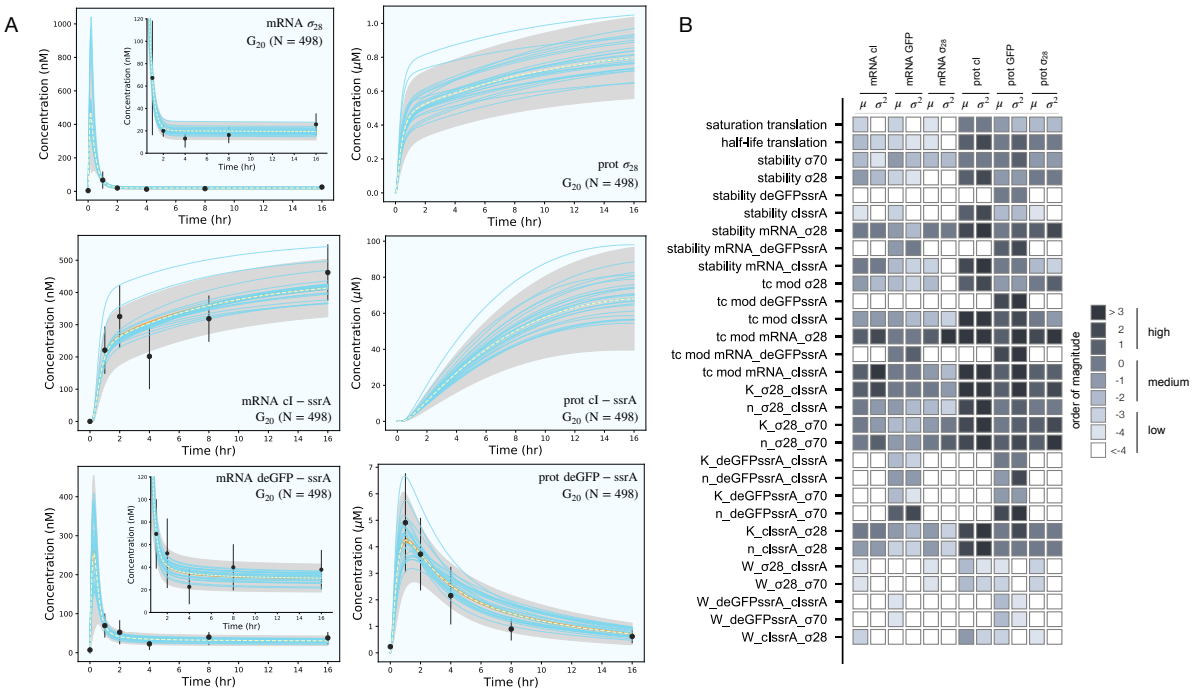

Fig. 3: Model simulations versus experimental measurements for the negative feedback deGFP-ssrA circuit. **A**: Model simulations of the negative feedback deGFP-ssrA circuit using the G_{20} ensemble (N = 498). Uncertainty: Simulations and uncertainty quantification are shown for the generation 20 (G_{20}) ensemble which yielded N = 489 parameter sets (rank two or below). The parameter ensemble was used to calculate the mean (dashed line) and the 99% confidence estimate of the simulation (gray region). Additionally, the 99% confidence estimate of the mean simulation is shown in orange. Individual parameter set trajectories are also shown in blue. Points denote the mean experimental measurement while error bars denote the 95% confidence estimate of the experimental mean computed from at least three replicates. **B**: Global sensitivity analysis of the negative feedback deGFP-ssrA circuit parameters. Morris sensitivity coefficients were calculated for the unknown model parameters, where the range for each parameter was established from the ensemble.

Table 1: Characteristic parameters for TX/TL model equations. Key to references used in the table: (a) [45], (b) [66], (c) set by experiment, (d) [67], (e) estimated in this study, (f) [68], (g) [69], (h) calculated from plasmid sequence, (i) [47], (f) [53]

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

Table 2: Estimated parameter values for the P70-deGFP model ($\mathcal{C}1$). The mean and standard deviation of each parameter value was calculated over the ensemble of parameter sets meeting the rank selection criteria (N = 139).

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

Table 3: Estimated parameter values for the negative feedback circuit ($\mathcal{C}2$). The mean and standard deviation for each parameter was calculated over the ensemble of parameter sets (N = 498).

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