Submitted, 2013 Conference on Decision and Control (CDC)
http://www.cds.caltech.edu/~murray/papers/tskm13-cdc.html

An *In Silico* Modeling Toolbox for Rapid Prototyping of Circuits in a Biomolecular "Breadboard" System

Zoltan A. Tuza[1], Vipul Singhal[2], Jongmin Kim[3], Richard M. Murray[3,4]

***Abstract*— In this paper, we develop an experimentally validated MATLAB software toolbox as an accompaniment to an *in vitro* cell-free biomolecular "breadboard" system. The toolbox gives insight into the dynamics of unmeasured states in the cell-free system, accounting especially for the resource usage. Parameter lumping and the reduced order modeling are used to maintain computational tractability and to avoid ill-conditioning. The toolbox allows for most applications to be implemented with standard set of commands for ease of use. Due to the breadboarding nature of the underlying cell-free system, the toolbox provides a general framework for experiment planning and predictive modeling for synthetic biomolecular circuits cell-free systems, accelerating our capacity to rationally design circuits from well characterized parts.**

I. INTRODUCTION

One of the major goals of synthetic biology is to apply a rational engineering design process to achieve the functioning of novel circuits *in vitro* and *in vivo* [15]. This bottom-up approach represents a shift towards design in a systematic and hierarchical way using well characterized and reusable biomolecular parts (DNA/plasmids, RNAs, enzymes, proteins, membranes) [5].

Even though many engineering methodologies and philosophies are being adopted by synthetic biologists, modeling remains challenging due to a variety of reasons, such as the existence of a large number of coupled (both known and unknown) mechanisms [3]. Furthermore, the boundaries of an artificial circuit introduced into an organism may be amorphous due to the interaction of the circuit with existing reactions in emergent ways. This, along with existing feedback regulation in organisms, leads to difficulties in troubleshooting synthetic circuits in an engineering sense.

These challenges are starting to be addressed by the paradigm of *in vitro* synthetic biology [6], especially through cell-free expression systems as a medium to build *in vitro* circuits. The cell-free expression system consists of crude cell extracts supplemented with buffer and resources [11], [22]. The crude cell extract contains functional protein machinery including transcription-translation machinery; however, all of the genetic material from host were carefully removed. Therefore, cell-free extracts provide a platform for the characterization of biomolecular parts and circuits in isolation, free from the natural regulation and the above mentioned "cross-talk" present in living cells [9]. A further advantage of cell-free *in-vitro* extracts is the significantly shorter characterization and design time cycles as compared to *in vivo* approaches. Despite recent progresses in streamlining the assembly and optimization process for synthetic circuits [16], achieving functional assembly for *in vivo* circuitry through conventional cloning poses significant challenges [14].

Here we present a software toolbox for MATLAB to accompany the TX-TL biomolecular breadboarding *in vitro* system developed by Shin and Noireaux [23]. This versatile *in vitro* system supports multiple stage cascades and bistable circuits as well as the expression of complete phage genome [21]. While elementary transcription and translation steps as well as degradation steps have been successfully modelled [12], the previous work focused on the linear regime and the initial stage of protein expression where the finite resources within the system are not limiting factors. In this work, by explicitly considering transcription, translation, and degradation machinery with energy sources, it is possible to capture experimental observations more quantitatively beyond the initial stage of experiment. This also provides insight into possible underlying reasons, such as resource limitations and enzyme loading [20], [19], [28], for previously unexplained behavior of the synthetic circuits being tested. With accurate modeling, the design cycle for prototyping may be shortened by eliminating circuit topologies or experimental conditions unlikely to meet design specifications. Moreover, predictive modeling is also possible with calibrated models, which may take the design process beyond the traditional trial-and-error approach. For these reasons, model-based circuit design and testing may be a valuable asset to synthetic biology.

This paper is organized as follows: In Section II we briefly outline the biomolecular breadboarding system and highlight its advantages and disadvantages over *in vivo* characterization and prototyping. In Section III we discuss the proposed mathematical model that captures the dynamics of the cell-free system, elaborating on the applied model reductions and parameter lumping. Section IV-A explains the applied system identification methodology. The experimentally validated examples are discussed in Section IV, where two simple circuits have been tested in the cell-free system and a set of parameters were estimated based on the experimental data. Finally, we state our conclusions and highlight future directions.

(1) Faculty of Information Technology, PPCU, H-1083 Budapest, Hungary tuza.zoltan@itk.ppke.hu
(2) Department of Computation and Neural Systems vsinghal@caltech.edu and (3) Department of Bioengineering and (4) Department of Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91006, USA

II. Cell-free system

The cell-free biomolecular "breadboard" system is a collection of *in vitro* protocols that can be used to test transcription and translation (TX-TL) circuits in a set of systematically-constructed environments that explore different elements of the external conditions in which the circuits must operate. This system is based on the work of Shin and Noireaux [23]. The transcription and translation machineries are extracted from *E. coli* cells. The endogenous DNA and mRNA from the cells are eliminated during the preparation. The resulting protein synthesis machinery is used to program cell-free TX-TL gene circuits in reactions. These gene circuits can be engineered in the laboratory using standard molecular cloning techniques, but it is also possible to use PCR products (linear DNA), which substantially decrease the design cycle time.

The cell-free *in vitro* system used here has certain advantages that make it desirable as a tool for simplifying the study of biological circuit function. Due to the lack of back regulation of most molecular species, it is possible to design and operate a synthetic circuit in different concentration regimes that may not be achievable in a living cell, for instance, due to copy number variability and incompatibility of plasmids. This in turn may allow us to explore a much larger parameter space to characterize the range of dynamical behaviors that a circuit topology may be capable of producing. Furthermore, the lack of back regulation and reaction for the cell usually translates to the same protocol being able to produce the circuit for different operating concentrations. Well characterized extracts and buffers, as in the case of the TX-TL system [23] also lead to a relatively accurate knowledge of the concentrations of the major components (nucleotides (NTPs), ribosomes, amino acids (AAs), RNA polymerases, among other things) of the system. It is also possible to characterize elementary dynamical processes like NTP consumption and the degradation of DNA and mRNA, leading to greater predictive capability of our computational model. Furthermore, in the cell-free system, there is negligible protein degradation [12], leading to longer lived reaction intermediates and products (such as fluorescent proteins). These proteins may, at the circuit designer's discretion, be tagged with 'degradation tags' at the time of gene design, allowing for degradation if desired.

On the other hand, the *in vitro* cell-free system also has some potential disadvantages over living cells, such as lack of growth, lack of sophisticated organization, and limited resources; though these do not necessarily pose a significant impediment to the prototyping function of the this breadboard system as shown by recent work [4]. The biggest caveat is that the *in vitro* cell-free system has finite resources (e.g., RNA polymerases, Ribosomes, NTPs, AAs), especially the loss of ATP has a strong impact on the system performance [13]. The second disadvantage of cell-free system is that there is no active growth nor waste removal process in the system, leading to the build up to reaction by-products such as non-functioning mRNA fragments and Adenosine diphosphate (ADP), which slow down reactions and eventually cause them to stop. The third difficulty is lack of compartmentalization and sophisticated organization on the surface. The finite resource and waste management problem can be partially addressed by an exchange dialysis system using membranes that allow exchange of fuel molecules and small wastes [23]. Furthermore, it is possible to implement compartmentalization and membranes in the form of tiny lipid vesicles in TX-TL [23]; however, the control of localization within the vesicle surface remains primitive. Highly controlled localization of molecular species to specific volumes is possible in cells, and may be desirable for certain circuit properties like ultrasensitivity [27].

III. Modeling approach

We have discussed the underlying *in vitro* biomolecular system in Section II, where we noted that a useful framework to study these systems is via the chemical reaction equations. The assumption of mass action kinetics (well-stirred and constant temperature mixture) allows these chemical equations to be written as a set of coupled ordinary differential equations. At the most general level, we can use the following deterministic system of ODEs to describe our model:

$$\begin{aligned} \dot{x} &= f(x,\theta) \\ y &= h(x,\theta) \end{aligned} \tag{1}$$

where $x \in R_+^n$ is the vector of modeled chemical species, $h : R^n \to R$ or R^2 is the output function, which is linear between the detection limits of instruments. The dynamics can be further described with the following factorization of the f vector field

$$f(x,\theta) = Nv(x,\theta) \tag{2}$$

where each row of the flux vector $v(x,\theta)$ corresponds to the rate at which a given reaction occurs and the corresponding column of the stoichiometric matrix N corresponds to the relative changes in concentration of the relevant species.

Before discussing how we modeled transcription and translation reactions, we briefly review the Michaelis-Menten kinetics used to model enzyme-mediated reactions: $E + S \overset{k_f,k_r}{\rightleftharpoons} E : S \xrightarrow{k_{cat}} E + P$, where the enzyme E and the substrate S bind to form an enzyme-substrate complex E:S, which in turn is converted into a product P and the enzyme E. We can find analytical expressions for the rate of production of P by making some simplifying assumptions: the $E : S$ complex is in a 'quasi–steady-state', with $(\dot{E : S}) = 0$ (Michaelis-Menten kinetics). This quasi–steady-state approximation is valid if $E^{tot} \ll S^{tot} + K_M$, where the total amount of E is constant $E_{tot} = E + E : S$, $S_{tot} = S + E : S$, and $K_M = (k_r + k_{cat})/k_f$. This leads to two regimes for the dependence of $\dot{P}$ on S_{tot}: when the amount of substrate is large, with $S_{tot} \gg K_M$, all of the enzyme is occupied by substrate and the rate of reaction does not depend on S_{tot}: $\dot{P} = k_{cat}E_{tot}$ (the zeroth order regime). When $S_{tot} \ll K_M$, it can be shown that the reaction rate is in the first order regime with respect to the amount

of substrate: $\dot{P} = (k_{cat}/K_M)E_{tot}S_{tot}$. These expressions give qualitatively accurate picture for the enzyme kinetics calculated by the more general mass-action kinetics in the modeling toolbox.

A. Transcription

Transcription, the process by which enzymes use nucleotide bases to create an RNA transcript from DNA, is modeled using the following chemical equations:

$$\mathrm{RNAP} + \sigma \longleftrightarrow \mathrm{RNAP} : \sigma \tag{3}$$

$$\mathrm{DNA} + \mathrm{RNAP} : \sigma \longleftrightarrow \mathrm{DNA} : \mathrm{RNAP} : \sigma$$

$$\mathrm{NTP} + \mathrm{DNA} : \mathrm{RNAP} : \sigma \longleftrightarrow \mathrm{NTP} : \mathrm{DNA} : \mathrm{RNAP} : \sigma$$

$$\mathrm{NTP} : \mathrm{DNA} : \mathrm{RNAP} : \sigma \longrightarrow \mathrm{DNA} + \mathrm{RNAP} : \sigma + \mathrm{mRNA}$$

The above equations describe transcription as a three step process: the activation of the core RNA polymerase (RNAP) by a sigma factor protein (σ) to form the holoenzyme (RNAP:σ); the binding of this activated RNAP to DNA to rapidly form a larger 'enzyme-template' complex DNA:RNAP:σ, which in turn uses NTP as a raw material to produce mRNA. For a typical transcription reaction utilizing the housekeeping sigma factor σ^{70}, the RNAP:σ will effectively be at an equilibrium after an initial transient, simplifying analysis as mentioned above. Furthermore, when other σ factors are utilized for transcriptional control (e.g., [23]), our approach can effectively capture the competition for core RNA polymerase (e.g., [24]). To further ensure computational tractability, we made a number of simplifying assumptions in the above transcriptional model: The four nucleotides (ATP, GTP, CTP and UTP) are lumped into one species called NTP, whose concentration is the sum of concentrations of the individual triphosphates. We obtain the estimates for RNAP and σ^{70} concentrations from a previous study [12] (these parameters can vary slightly depending on the batch); the NTP concentrations are determined from buffer preparation ([ATP]=[GFP]=1.5 mM, [CTP]=[UTP]=0.9 mM) [26]. We do not model initiation and elongation separately, instead simply model the production of the whole mRNA transcript in a single enzymatic step (A more detailed model of mRNA production including competitive inhibition can be found in [1]). To avoid numerical issues and keep track of material balance (stoichiometry), we also implemented a 'NTP consumption' reaction, whose rates are pegged with the transcription rate and the length of transcript.

There are two important determinants of the speed of transcription: the concentration of DNA templates and the concentration of NTPs. At low DNA template concentrations, transcription rate is approximately linear with respect to the template concentration (first-order reaction), while at high DNA template concentrations, the transcription rate is almost independent of DNA concentration. As previously mentioned, this regime is called the zeroth-order regime, and is believed to be due to the saturation of the transcriptional machinery. The existence of a saturation regime in the cell-free expression system has been previously demonstrated in the context of protein outputs [23]. During the initial phase of transcriptional reactions, NTPs are in excess of their Michaelis constants, approximating zeroth-order Michaelis-Menten kinetics with respect to the NTP concentration; however, after a significant portion of the NTP pool has been consumed, the binding of the NTPs to the enzyme-template becomes one of the rate-limiting factors, approximating first-order Michaelis-Menten kinetics with respect to NTP concentration: ($[\dot{mRNA}] \propto [NTP]$).

The cell extract also contains the machinery for mRNA degradation (presumably several endonucleases and exonucleases remain active in cell-free extract), which has been studied in a previous work [12]. Phenomenologically, the mRNA degradation followed a first-order kinetics with a half life of 12 minutes. Thus, we modeled mRNA degradation to be a first order process as follows (the enzyme-substrate complex is not modeled):

$$\mathrm{mRNA} + \mathrm{RNase} \longrightarrow \mathrm{RNase} \tag{4}$$

Finally, the transcription reaction slows down and eventually stops because the NTP levels became too low or the build up of adenosine diphosphate (ADP) relative to ATP makes enzyme reactions energetically unfavorable [13]. Instead of modeling ATPs versus other nucleotides or their phosphorylation states separately, we model the decrease in ATP usability in a lumped manner through the first order degradation of NTP. This decrease in NTP concentration leads to a fall in the rate of mRNA production to below that of mRNA degradation, causing the mRNA concentration to slowly drop to zero.

B. Translation

The transcribed messenger RNA is then used as a template to create a nascent protein via translation. In detailed models, a ribosome (Ribo) sits on the mRNA transcript and the charged transfer RNA (tRNA) molecules are used to transport specific amino acids to the ribosome:mRNA complex to elongate the growing polypeptide chain. The specific tRNA that binds to the elongation site, and hence the amino acid incorporated, depends on the nucleotide triplet (codon) on the mRNA being read. The endogenous enzyme, aminoacyl-tRNA-synthetase, charges the tRNA molecules with amino acids; however, the concentration of endogenous aminoacyl-tRNA-synthetase is uncertain. The endogenous tRNA concentration in the cell-free expression is also uncertain; we add 8 μM of additional tRNA in the buffer to ensure that there is sufficient concentrations of all tRNAs in the mixture. Furthermore, we ensure that there is an excess of AA in the mixture, so that tRNAs can be assumed to be constantly charged by AAs until the concentration of AAs falls close to those of tRNAs (on the order of 10 μM). Thus, we approximate the saturation regime of tRNA-AA through a reaction where a single species AA binds to the ribosome-mRNA complex with the binding constant set close to total tRNA levels (on the order of 10 μM). The rate of this reaction

lumps the time required for tRNA charging and the transport of tRNA to the ribosome by diffusion. In all, we model translation as follows:

$$\mathrm{mRNA} + \mathrm{Ribo} \longleftrightarrow \mathrm{mRNA} : \mathrm{Ribo} \tag{5}$$

$$\mathrm{AA} + \mathrm{mRNA} : \mathrm{Ribo} \longleftrightarrow \mathrm{AA} : \mathrm{mRNA} : \mathrm{Ribo}$$

$$\mathrm{AA} : \mathrm{mRNA} : \mathrm{Ribo} \longrightarrow \mathrm{mRNA} + \mathrm{Ribo} + \mathrm{Protein}$$

Here, we have lumped initiation, elongation, and termination into a single reaction of AA binding to the ribosome-mRNA complex leading to protein production and the dissociation of the complex. To avoid numerical issues and to keep track of material balance (stoichiometry), we also implemented an 'AA consumption' reaction, whose rate depends on the translation rate and the length of protein. The energy requirement of translation is not modelled, and this is partially justified because ATP and GTP are provided in excess of CTP and UTP in anticipation of the additional usage for translation (about one ATP and one GTP are used for each amino acid incorporated). An explicit resource usage model for keeping track of each NTP separately will be explored in future work. Because the degradation of protein is negligible without degradation tag, the protein concentration reaches a constant level only when the mRNA runs out.

C. Enzyme loading

One set of hidden dynamics that the modeling toolbox provides insight into is enzyme loading effects. In most reactions, RNAP saturates at low nM range of DNA templates with strong promoters. As the initial gene concentrations are varied, or multiple genes are expressed in the modeling toolbox, the ribosomes can be either in an approximately linear regime or in saturation, and this leads to different dynamics. We are in the process of characterizing these effects.

D. Protein degradation

As it was mentioned in section III-B, no significant protein degradation has been observed in the TX-TL system. Protein degradation requires a degradation tag on the target protein which is recognized by a specific ClpX:ClpP protease complex. ClpX unfolds tagged protein using ATP, followed by ClpP degrading this unfolded protein into nonfunctional constituents. Endogenous ClpX:ClpP complex in the TX-TL extract is capable of degrading $0.5\mu M$ tagged protein. For this reason, extensive protein degradation requires the expression of ClpX and ClpP proteins within the system. From the modeling point of view, the mehanism of the action of the proteases is very complex. The energy usage of ClpX cannot be neglected because the degradation stops when energy source (NTP) levels drop in the solution [2]. Moreover, a recent study suggested that ClpP enhances the success rate of ClpX, while ClpX becomes non-functional after a certain number of unfolding attempts. Thus, our preliminary model is:

$$\mathrm{P_{SSRA}} + \mathrm{ClpX} \longleftrightarrow \mathrm{P_{SSRA}} : \mathrm{ClpX} \tag{6}$$

$$\mathrm{NTP} + \mathrm{P_{SSRA}} : \mathrm{ClpX} \longleftrightarrow \mathrm{NTP} : \mathrm{P_{SSRA}} : \mathrm{ClpX}$$

$$\mathrm{NTP} : \mathrm{P_{SSRA}} : \mathrm{ClpX} \longrightarrow \mathrm{P^{-}_{SSRA}}$$

$$\mathrm{P^{-}_{SSRA}} + \mathrm{ClpP} \longleftrightarrow \mathrm{P^{-}_{SSRA}} : \mathrm{ClpP}$$

$$\mathrm{P^{-}_{SSRA}} : \mathrm{ClpP} \longrightarrow \mathrm{ClpP}$$

$$\mathrm{ClpX} \longrightarrow 0$$

We assume that concentration ClpP is already enhancing the ClpX activity since we have added the same amount of ClpX and ClpP plasmids. We are planning to give experimental results and system ID results on these in the future work.

E. Linear DNA protection

Protein production from a linear DNA is also possible in the *in-vivo* system, reducing the design cycle time significantly. Linear DNA, unlike plasmids, is degraded by endogenous RecBCD. The degradation can be attenuated significantly with the addition of purified gamS protein. This protein binds to RecBCD with a high affinity and sequesters it away form the DNA. In our toolbox this process is modeled as follows:

$$\mathrm{RecBCD} + \mathrm{gamS} \longleftrightarrow \mathrm{RecBCD} : \mathrm{gamS} \tag{7}$$

$$\mathrm{DNA} + \mathrm{RecBCD} \longleftrightarrow \mathrm{DNA} : \mathrm{RecBCD}$$

$$\mathrm{DNA} + \mathrm{RecBCD} \longrightarrow \mathrm{RecBCD}$$

The DNA:RecBCD complex has a larger dissociation constant than the RecBCD:gamS complex [18]. Thus when gamS is present in excess, the DNA degradation almost stops.

F. Implementation considerations

Both transcription and translation are implemented as one step reactions, which, under a naive implementation scheme:

$$\mathrm{n{\cdot}NTP} + \mathrm{DNA} : \mathrm{RNAP} : \sigma \longleftrightarrow \mathrm{n{\cdot}NTP} : \mathrm{DNA} : \mathrm{RNAP} : \sigma \tag{8}$$

$$\mathrm{n{\cdot}NTP} : \mathrm{DNA} : \mathrm{RNAP} : \sigma \longrightarrow \mathrm{DNA} + \mathrm{RNAP} : \sigma + \mathrm{mRNA}$$

with a large stoichiometric coefficient n in front of the NTP (AA) term in the chemical equation. This leads to numerical difficulties: 1) the ODEs describing these reactions become very stiff due to an extremely high reaction order and 2) the reaction order and hence rate depends on the length of the mRNA, because n is simply this length in nucleotide bases (this is not an accurate model, since the whole mRNA is not assembled instantaneously, like it suggests).

We first alleviated these problems by clumping the NTP (AA) to be consumed in units of 100, bringing the nominal reaction order down to 6-12 (the genes are typically 600-1200 bases long). In the modeling toolbox, we used a heuristic approach whereby we used only one unit of NTP (AA) for each transcription (translation) reaction and coupled this to the implementation of a "dummy" reaction to consume the rest.

IV. Experimental validation

In the remainder of the paper we will characterize the unknown reaction rates and demonstrate the capabilities of the modeling toolbox. To this end, we attempted to identify transcription and translation rates individually. mRNA dynamics were measured with a spectrofluorometer and protein concentrations were measured with a plate reader. Finally, we will illustrate the regulatory capabilities of the *in vitro* cell-free system and compare it with the results from the computational model.

A. System identification

The core TX-TL model contains 11 reactions and 11 parameters in total, along with 7 species with non-zero the initial concentration (e.g. enzymes, amino acids, and nucleotides). Some of the parameters are known: either they are part of the assembly protocol [26] (e.g. NTP, DNA concentration) or they were studied previously (e.g. concentration of RNA polymerase and ribosome, mRNA degradation rate, etc) [23]. The rest of the parameters were subject to parameter estimation (two for transcription and two for translation).

From (1) we can see that the model is nonlinear, thus it is difficult to find an analytical expression for parameter values. Traditional random search methods provide only point estimates [10], whereas other methods, one of which is used below, can sample the distribution of unknown parameters. The result is that we obtain not only point estimates, but also statistics about individual parameters and correlations between them as well [8].

Using the simplex search algorithm with quadratic cost function, we found a suitable start point for the Markov Chain Monte Carlo (MCMC) algorithm to sample the joint distribution. In each case the cost function is given by:

$$J(\theta) = e(\theta)^T Q e(\theta) \tag{9}$$

where θ is the parameter vector, $e(\theta)$ is the difference between the measured output trajectory y and the model output $y(\theta)$ for a given θ. In our case, the weighting matrix Q was the identity.

In our reduced order model, the whole gene expression process can be viewed as two enzymatic conversions: first NTP is converted to mRNA, then AA converted to unfolded protein. In translation, mRNA is part of the enzyme complex, and therefore its concentration changing over time. In the first step, we assume that RNAP and the sigma factor forms a complex quickly with very high disassociation constant (K_d 10 μM). Then the $RNAP\sigma : DNA$ complex is formed with much lower disassociation constant (K_d 5 nM).

For transcription, we have determined two parameters: the degradation rate of NTP (all nucleotides are lumped into a specie named NTP) and the Michaelis Menten constant ($K_m = (k_r + k_{cat})/k_f$) for the $RNAP : \sigma$:DNA enzyme-template. Since the rate of transcription (k_{cat}) is known [12], we can fix the forward rate k_f to find the reverse rate k_r). Using the two step estimation process, we determined the ATP degradation rate to be $9.32 \times 10e^{-4}\ s^{-1}$. (confidence interval $[8.2e-04\ \ 10e-4]$) and the K_m to be $6.92\mu M$ (confidence interval [5.6744 8.1656])

For translation, the dissociation constant $k_d = k_r/k_f$ for the mRNA:Ribo complex and the K_m for the mRNA:Ribo "enzyme-template have been determined. Again, the k_{cat} rate is known from a previous study [12], we are only searching for the reverse rate. After running the two step algorithm, we found the $k_d = 0.0500$ ([0.0445 0.0555]) and the $K_m = 168\mu M$ ([132.72 203.28]).

In the rest of the paper, these parameters are used for simulations and also as a starting point for biomolecular circuits (e.g. the negative autoregulation example in Section IV-E).

B. Experimental procedure

The cell extract is a crude cytoplasmic extract from E. coli which contains soluble proteins, including the entire endogenous TX-TL machinery, as well as mRNA and protein degradation enzymes [22], [23]. (Detailed instructions on the cell extract preparation can be found in [26]). Reactions typically take place in volumes of 10 μl (with a plate reader) to 60 μl (with a spectrofluorometer) at 29°C. mRNA dynamics can be followed using an RNA aptamer for the fluorescent dye malachite green recorded in spectrofluorometer (FluoroLog-3, Horiba Jobin Yvon); protein dynamics were followed by fluorescence of GFP recorded in the plate reader (Victor X3, PerkinElmer). Two variants of GFP were used: deGFP, eGFP variants optimized for *in vitro* experiments [22], and deGFP fused to repressor tetR. To avoid variation between extract batches we used the same extract for all experiments.

C. mRNA dynamics

Radiolabeling of mRNA and subsequent gel analysis can provide limited temporal resolution for mRNA dynamics [12]. Therefore, we performed real-time fluorescence monitoring of mRNA dynamics by utilizing an RNA aptamer with the fluorescent dye malachite green [7]. The aptamer contains a binding pocket for the malachite green dye but is very short (35 bases), and thus, the aptamer sequence was incorporated at the 5′ UTR of the plasmid encoding *deGFP* gene. The fluorescence signal from this transcript is comparable to the short malachite green aptamer itself, but the half-life in cell-free extract is much longer for the plasmid-aptamer-malachite green complex. To determine mRNA dynamics at high sensitivity with high temporal resolution, we used a spectrofluorometer with the following setting: the excitation and emission maxima were set at 630 nm and 655 nm with 5 nm slit width using monochromater with fluorescence measured every minute. Upon the initiation of reactions by moving the sample to 29°C chamber, the RNA level quickly increased to its peak level at approximately 60 minutes after the start of the experiment, and slowly decayed over time. Note that the RNA level varied linearly with respect to DNA template concentrations (up to 3 nM that we tested). (In principle, other promoters with weaker promoter strengths can be used for probing mRNA dynamics in a similar manner. However, the signal-to-noise ratio for

the measurements when the Pr1 and Pr2 promoters were used (see below), was low even with the spectrofluorometer.) Compared to a previous modeling study [12], we observed that the initial delay in the accumulation of mRNA was smaller and the apparent peak for mRNA level was higher. These differences may be explained by the fact that the malachite green aptamer is located at the 5′ UTR of mRNA for deGFP: the beginning of transcription would lead to immediate increase in fluorescence signal while the decay of transcript could be reported at a later time than the initial degradation by endonuclease.

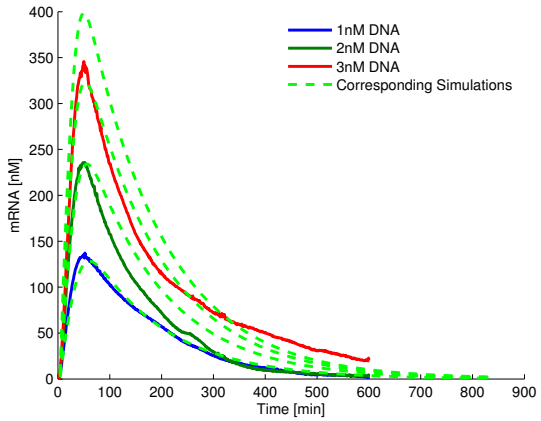

Fig. 1. mRNA dynamics were measured for 10 hours via fluorescent RNA aptamer monitoring with 3 different plasmid concentrations containing the same circuit. Then we used that data for parameter estimation, the result of that is shown in green, where we simulated the mRNA production with 4 different concentration.

D. protein production with constitutive promoter

One of the simplest reactions to be modeled is the constitutive production of a reporter protein. The general structure of the DNA looks like promoter-UTR-gene. With proper optimization of reaction conditions, the TX-TL *in vitro* system (relying exclusively upon endogenous *E. coli* RNAP) is shown to accumulate 0.5 to 1 mg/mL of reporter protein (plate reader fluorosence measurement. Excitation: 485nm, Emission: 525nm, Bandwidth 30nm, Measurements every 3 minutes), comparable to other *in vitro* cell-free expression systems that use bacteriophage RNA polymerase [11]. Because endogenous proteases in the TX-TL system are in a small quantity and typically require degradation tags for their action [12], degradation on untagged reporter proteins is negligible. We used three constitutive promoters with different promoter strengths: Pr, Pr1, and Pr2. Pr, the lambda repressor Cro promoter, is the strongest of these, while mutants Pr1 and Pr2 have activities of 20% and 2% of the activity of Pr. As expected, the reporter deGFP protein accumulates fastest under the Pr promoter and slowest under the Pr2 promoter. Figure 2 and 3 show the activity of the two promoters. The protein concentration scales linearly with initial plasmid concentration up to 3 nM for the strongest promoter, beyond which the scaling is nonlinear. It remains to be verified that the transcription machinery (RNAP:σ complex is not saturated by DNA at the 1 nM to 3 nM plasmid concentration and thus operates in the first order regime. Beyond 3nM of plasmid DNA concentration, the machinery may start to get saturated, leading to nonlinear scaling.

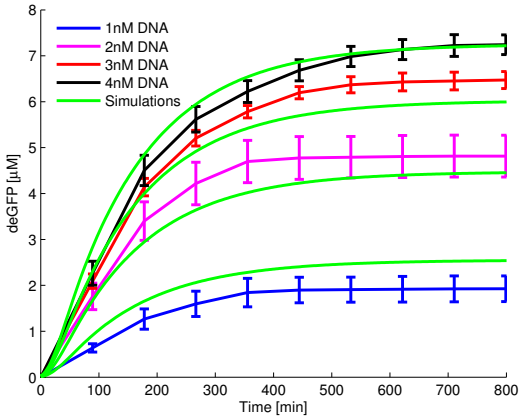

Fig. 2. Figure shows the constitutive production of reporter protein under the strongest promoter. The initial plasmid concentration was varied between 1nM and 4nM ± 5%, using estimated parameters from IV-A. Simulations with varying plasmid concentration were run, and the results are shown in green for each case. Beyond the 3 nM case, the steady-state value of the reporter protein does not scale linearly with the initial plasmid concentration

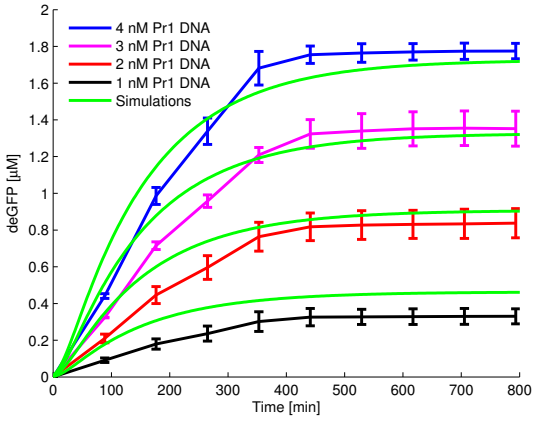

Fig. 3. The constitutive production of reporter protein under the second strongest promoter (Pr1) was also tested. We varied the initial plasmid concentration between 1nM and 4nM ± 5%. The steady-state value of the reporter protein scales linearly throughout the measured region. The reader should note that the activity of 4nM± 5% of this the plasmid is comparable to the activity of the strongest promoter.

E. Negative Autoregulation

An approach to testing the regulatory potential of the TX-TL system, is to use repressors not present in the *E. coli* extract. To this end, we utilized the tetracycline system, constructed as a negative feedback loop, with the synthetic regulatory element PLtetO$_1$, composed of a strong promoter specific to σ^{70} bound RNAP and two tet operators [17].

The tetracycline repressor gene *tetR* and the *deGFP* gene were fused and cloned under $\mathrm{PLtetO_1}$ into one plasmid. To determine reporter deGFP protein dynamics, we used plate reader with the same setting in the previous section. Repression of deGFP expression is observed after 30 minutes of incubation in the absence of the inducer anhydrotetracycline (aTc) (Figure 4 shown in red) (measurement as in section IV-D). On the other hand, the reporter deGFP is fully expressed when a concentration of 10 μM (5 μg/mL) of aTc is included in the reaction. The high concentration range of aTc required for full expression, about 50 fold higher than the amount used for in vivo induction [17], is in agreement with the previous analysis of an analogous tetracycline circuit without deGFP fusion protein [23].

We modeled the promoter activity of $\mathrm{PLtetO_1}$ as follows (see also [25]):

- tetR repressor protein, produced as a monomer, binds to another monomer to form a tetR dimer ($K_d = 10^{-7}$ to 10^{-8} M);
- tetR dimer binds to the promoter $\mathrm{PLtetO_1}$ and blocks RNAP binding ($K_d = 10^{-12}$ to 10^{-13} M); the two operator regions within the promoter were not separately modeled;
- aTc binds to tetR monomer and tetR dimer (irrespective of its binding state to promoter); tetR bound to aTc binds to the promoter weakly ($K_d = 10^{-5}$ to $K_d = 10^{-6}$ M).

Following the previous notation, the above processes can be modeled with these additional reactions:

$$\mathrm{tetR} + \mathrm{tetR} \longleftrightarrow \mathrm{tetRdimer} \tag{10}$$

$$\mathrm{DNA} + \mathrm{tetRdimer} \longleftrightarrow \mathrm{DNA} : \mathrm{tetRdimer}$$

$$2\,\mathrm{aTc} + \mathrm{tetRdimer} \longleftrightarrow 2\,\mathrm{aTc} : \mathrm{tetRdimer}$$

Using the parameter estimation technique introduced above, we estimated the K_m for the promoter using the previously identified values.

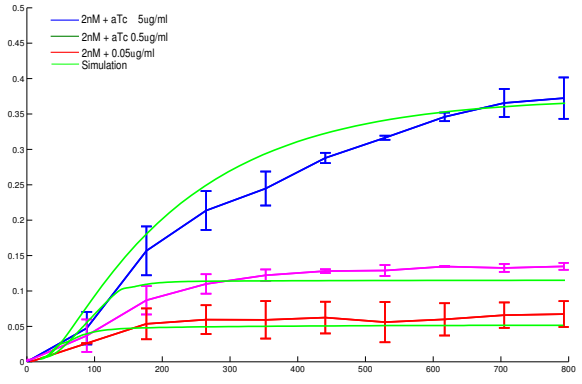

Fig. 4. In this experiment using fixed concentration of 2nM PLtetO-UTR-deGFP/tetR plasmid and varying amount of anhydrotetracycline, we measured the reporter protein concentration with three different levels of aTc. With green, we show the corresponding simulation results.

V. CONCLUSIONS

We have demonstrated a reduced order model for our cell-free system, where the modeled states captures the observable dynamics with minimal model complexity. Because of this rapid prototyping is possible, where multiple circuit topologies and corresponding parametrizations can be tested at once. The developed model also has predictive capability because the parameters have been validated experimentally. Furthermore, the model provides insight into unmeasured states, which enables the user to keep track of resource consumption and enzyme loading.

VI. ACKNOWLEDGMENTS

This work is supported by the DARPA Living Foundries Program under Contract HR0011-12-C-0065 and by TAMOP-4.2.1-B-11/2/KMR-2011-0002. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

[1] Sabine Arnold, Martin Siemann, Kai Scharnweber, Markus Werner, Sandra Baumann, and Matthias Reuss. Kinetic modeling and simulation of *in vitro* transcription by phage T7RNA polymerase. Biotechnology and Bioengineering, 72:548–561, 2001.

[2] Tania A. Baker and Robert T. Sauer. ClpXP, an ATP-powered unfolding and protein-degradation machine. Biochimica et Biophysica Acta, 1823:15–28, 2012.

[3] Stefano Cardinale and Adam Paul Arkin. Contextualizing context for synthetic biology identifying causes of failure of synthetic biological systems. Biotechnology Journal, 7:856–866, 2012.

[4] James Chappell, Kirsten Jensen, and Paul S. Freemont. Validation of an entirely *in vitro* approach for rapid prototyping of dna regulatory elements for synthetic biology. Nucleic Acids Research, 2013.

[5] Drew Endy. Foundations for engineering biology. Nature, 438:449–453, 2005.

[6] Anthony C. Forster and George M. Church. Synthetic biology projects *in vitro*. Genome Research, 17:1–6, 2007.

[7] Dilara Grate and Charles Wilson. Laser-mediated, site-specific inactivation of RNA transcripts. Proc Natl Acad Sci USA, 96:6131–6136, 1999.

[8] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. Bernoulli, 7:223–242, 2001.

[9] C. E. Hodgman and M. C. Jewett. Cell-free synthetic biology: thinking outside the cell. Metab. Eng., 14(3):261–269, 2012.

[10] Lagarias J.C., J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. SIAM Journal of Optimization, 9:112–147, 1998.

[11] Michael C Jewett, Kara A Calhoun, Alexei Voloshin, Jessica J Wuu, and James R Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. Molecular Systems Biology, 4:220, 2008.

[12] Eyal Karzbrun, Jonghyeon Shin, Roy H. Bar-Ziv, and Vincent Noireaux. Coarse-grained dynamics of protein synthesis in a cell-free system. Phys. Rev. Lett., 2011.

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

[14] Roberta Kwok. Five hard truths for synthetic biology. Nature, 463:288–290, 2010.

[15] Michael Elowitzand Wendell A. Lim. Build life to understand it. Nature, 468:889–890, 2010.

[16] Kevin D Litcofsky, Raffi B Afeyan, Russell J Krom, Ahmad S Khalil, and James J Collins. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. Nature Methods, 9:1077–1080, 2012.

[17] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in *escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Research, 25:1203–1210, 1997.

[18] Kenan C. Murphy. The lambda Gam protein inhibits RecBCD binding to dsDNA ends. Journal of Molecular Biology, 371:19–24, 2007.

[19] Yannick Rondelez. Competition for catalytic resources alters biological network dynamics. Phys Rev Lett, 108(1):018102, 2012.

[20] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. Science, 330(6007):1099–1102, 2010.

[21] Jonghyeon Shin, Paul Jardine, and Vincent Noireaux. Genome replication, synthesis and assembly of the bacteriophage T7 in a single cell-free reaction. ACS Synth. Biol., 1:408413, 2012.

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

[23] Jonghyeon Shin and Vincent Noireaux. An E. coli cell-free expression toolbox: Application to synthetic gene circuits and artificial cells. ACS Synth. Biol., 1:29–41, 2012.

[24] Dan Siegal-Gaskins, Vincent Noireaux, and Richard M. Murray. Biomolecular resource utilization in elementary cell-free gene circuits. In Proceedings of the American Control Conference, 2013.

[25] Vassilios Sotiropoulos and Yiannis N Kaznessis. Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. BMC Systems Biology, 1:7, 2007.

[26] Zachary Z. Sun, Clarmyra A. Hayes, Jonghyeon Shin, Filippo Caschera, Richard M. Murray, and Vincent Noireaux. Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. Journal of Visualized Experiments, 2013.

[27] Siebe B van Albada and Pieter Rein ten Wolde. Enzyme localization can drastically affect signal amplification in signal transduction pathways. PLoS Comput Biol., 3, 2007.

[28] Enoch Yeung, Jongmin Kim, Ye Yuan, Jorge Gonçalves, and Richard Murray. Quantifying crosstalk in biochemical systems. In Proceedings of the IEEE Conference on Decision and Control, 2012.