Submitted, 2014 Winter q-bio Conference (5 Nov 2013)
http://www.cds.caltech.edu/~murray/papers/ygm14-wqbio.html

System identification of the phosophorylation based insulator in an *in vitro* cell-free expression system

Enoch Yeung, Shaobin Guo, Richard M. Murray

Abstract

An outstanding challenge in the design of synthetic biocircuits is the development of a robust and efficient strategy for interconnecting functional modules. Recent theoretical work demonstrated that a phosphorylation based insulator implementing a dual strategy of high gain and strong negative feedback could potentially serve as a device to attenuate retroactivity. This research investigates the structural identifiability of the phoshorylation based insulator when implemented in a transcription-translation (TXTL) cell free expression system. We consider a complex model that provides an intricate description of all chemical reactions and leveraging specific physiologically plausible assumptions, we derive a rigorous simplified model that captures the output dynamics of the phosphorylation based insulator. We perform standard system identification analysis and determine that the model is globally identifiable with respect to three critical parameters. These three parameters are identifiable under specific experimental conditions and we perform these experiments to estimate the parameters. Our experimental results suggest that the functional form of our simplified model is sufficient to describe reporter dynamics and enable parameter estimation. In general, this research illustrates the utility of the TXTL cell free expression system as a platform for system identification, as it provides extra control inputs for parameter estimation that typically are unavailable in vivo.

I. Introduction

One grand challenge in synthetic biology is understanding how functional modules can be efficiently and robustly interconnected to yield more complex systems. Despite the development of *de novo* synthetic biological modules capable of diverse functions, e.g. RNA-based transcriptional regulation [1], rewritable digital storage [2], multilayered logic [3], and robust oscillations [4], an experimental framework for integrating such modules has yet to be demonstrated.

A difficulty that arises upon the interconnection of two modules in series is a phenomenon called retroactivity [5], [6]. Even if one of the modules is designed to be upstream or driving the input of the second module, a retroactive signal from the downstream (second) module is generated upon interconnection of the two modules that alters the internal dynamics of the first module. In [5], the authors show that by inserting an additional insulator module, using a strategy of negative feedback coupled with high gain, the insulator can effectively attenuate retroactivity. In particular, they illustrated the conceptual design of a phosphorylation based insulator: negative feedback on the output of the first module is obtained by toggling the output from an active or phosphorylated state to a dephosphorylated state and high gain is implemented by tuning the amount of kinase produced in the system. An *in vivo* implementation of this phosphorylation based insulator in *S. cerevisiae* is the subject of ongoing work.

On the theoretical side, there have been significant discoveries on the subject of retroactivity. In [5] the authors developed a rigorous mathematical definition for describing and quantifying retroactivity between two modules. In [7] formal expressions for intramodular and intermodular retroactivity are derived for complex gene transcription networks consisting of nodes, modules, and systems. Additionally, the authors in [8] show that long signal cascades can attenuate retroactivity, while [9] utilize a time-scale separation strategy to attenuate retroactivity.

Finally, in [10] the authors show that attenuating retroactivity can have the unwanted effect of amplifying high frequency noise in gene expression. The pace at which these theoretical results emerge is often much faster than the rate at which experimental implementations of these theoretical concepts are achieved.

Technologies that allow for rapid exploration or validation of theoretical predictions are desirable. However, a major experimental obstacle in rapidly developing insulators to attenuate retroactivity is the optimization of multiple design variables. In the phosphorlyation based insulator, the copy number of the kinases and phosphatases must be tuned to achieve an appropriate gain. The copy number of these two enzymes can be modified by an appropriate choice of plasmid replication origin (or gene copy number in the case of chromosomal integration), ribosome binding site, and promoter type. However, this leads to a combinatorial number of realizations of the phosphorylation based insulator. Using state of the art cloning techniques, each realization requires at least two to three days to synthesize [11]. Thus, it would be extremely valuable to have a prototyping platform or breadboard system for rapidly exploring the realization space of all possible implementations of the phosphorylation based insulator.

The cell free *in vitro* transcription translation (TXTL) system developed in [11], [12] is an attractive candidate platform for such rapid protoyping. The system facilitates DNA based expression on plasmids and linear DNA [13] and since linear and plasmid DNA can be synthesized and expressed in the TXTL system in a single day's time, the time required to iterate over designs is considerably reduced.

Another powerful aspect of the TXTL system is the ability to directly modulate the concentration of different pieces of DNA encoding different biocircuit components. The ability to rapidly synthesize and test the effect of different promoter sites, ribosome binding sites, untranslated regions (UTR), etc. and simultaneously vary the DNA encoding these parts permits a degree of freedom typically absent in *in vivo* assays. In this setting, iterating of prototypes could be assisted by predictive modeling of biocircuit dynamics. It is the ability to control DNA concentrations and rapidly vary structural properties of the biocircuit that allow us to address the problem of parameterizing a predictive model.

In vitro systems have long been used to characterize fundamental parameters in biological systems []. In a synthetic biology context, especially for the phosphorylation based insulator circuit, it is unclear what parametric information can be extracted from a series of systematic tests in an *in vitro* system, specifically the TXTL system. With additional degrees of freedom in the experimental conditions, the TXTL system may be able to provide insight into model parameters that *in vivo* studies could not. Moreover, it is unclear what systematic tests should be carried out in order to tease out this information. This paper investigates these issues using the phosphorylation based insulator as a case study.

In general, a parametric model is globally structurally identifiable only under certain mathematical conditions [14]. These conditions are valid as long as the control variables enter the dynamical system as a multiplicative perturbation. However, as we will see with the phosphorylation based insulator, even if the model retains this structure the model may not be globally identifiable because of the large number of parameters it contains, despite having only a couple output variables. As is often the case, a first principles model may be physically representative of the intricate reactions happening in the system, but carry a complexity that far exceeds the information present in the data. Thus, simplified models that are reflective of the low-dimensional output data, while also retaining the (controllable) experimental variables in the TXTL system are desirable.

Therefore, in this work we propose a complex model based on the fundamental processes of transcription, translation, and phosphorylation. The model is unwieldy to analyze and so we rigorously derive a simplified

model based on a series of physically realistic assumptions, show that it is globally identifiable with respect to the data and perform a series of experimental perturbation tests to back out the simplified model parameters.

II. Background: The Phosphorylation Based Insulator

The phosphorylation based insulator is a biocircuit designed to insulate two biocircuit modules connected in series from the effects of retroactivity. An upstream module S_1 has input u_1 and output y_1 while the downstream module S_2 has input u_2 and output y_2. In a system free of retroactivity effects, the signal $u_2 = y_1$ and there is no signal that maps from S_2 to S_1. However, in biological systems, the output of a system is often a molecule that is consumed or incorporated in a downstream process. In terms of our model, once S_2 is interconnected to S_1, a retroactive input r from S_2 to S_1 that modifies the dynamics of the output y_1 must be incorporated in our model. The variable r is referred to as retroactivity to the output y_1 and also retroactivity to the input u_2. A general framework is derived in [10] to describe arbitrary systems, but for our analysis of the phosphorylation based insulator, we will suppose in the native or uninsulated system we have two modules S_1 and S_2 where the dynamics S_1 are given as:

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, u_1, r) \\ y_1 &= Y_1(x_1, u_1, r); \end{aligned} \tag{1}$$

and the dynamics of S_2 are given as

$$\begin{aligned} \dot{x}_2 &= f_2(x_2, u_2) = f_2(x_2, y_1) \\ y_2 &= Y_2(x_2, u_2) = Y_2(x_2, y_1) \\ r &= R(x_2, u_2) \end{aligned} \tag{2}$$

Notice that due to the nature of interconnections in biology, the states x_1 of S_1 may have overlap with the inputs u_2 of S_2. In particular, even though y_1 may also be a state in S_1, $y_1 = u_2$ is the input for S_2. Additionally, we will consider systems S_1 and S_2 to be interconnected only to each other. Thus, those familiar with the work of [10] will note that we have omitted the retroactivity to the input signal for S_1 and the retroactivity to the output signal of S_2. The retroactivity signal we wish to center our attention on is r, which describes the retroactivity to the output signal for S_1, or viewed from the perspective of S_2, the retroactivity to the input signal for S_2. In Figure 1 we provide a schematic illustrating the interconnection of the two systems and their respective signals.

The purpose of the phosphorylation based insulator is to insulate systems S_1 and S_2 from the retroactivity effects that arise from their interconnection. Mathematically, the insulator can be considered as a separate system S_I with the express purpose of eliminating retroactivity to the output y_1 (and equivalently retroactivity to the input u_2). In the literature, the insulator is inserted as a separate system in between S_1 and S_2 — we will adopt the same paradigm. Thus, the dynamics of the insulated system are given as

$$\begin{aligned} \dot{x}_1 &= f_1(x_1, u_1) \\ y_1 &= Y_1(x_1, u_1); \end{aligned} \tag{3}$$

and the dynamics of S_I (denoted S_{Ins} in Figure 1) are given as

$$\begin{aligned} \dot{x}_I &= f_I(x_I, u_I, r) \\ y_I &= Y_I(x_I, u_I, r) \end{aligned} \tag{4}$$

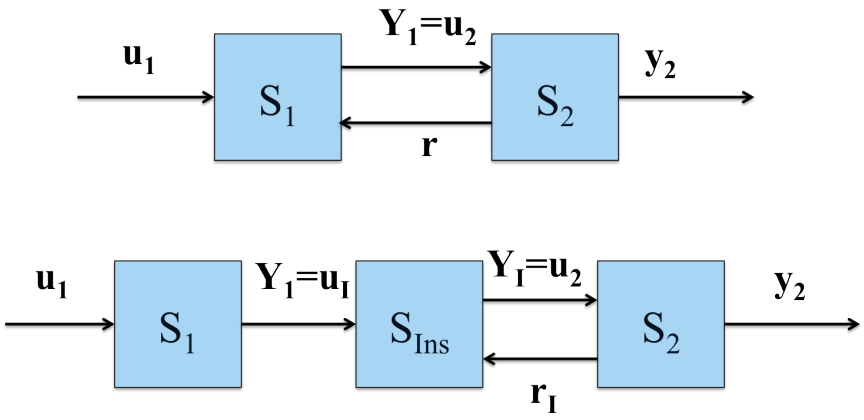

Fig. 1: A schematic illustrating the structure and function of the phosphorylation based insulator. In a natural uninsulated setting, upon interconnection of an upstream system (S_1) and downstream system (S_2), a retroactivity signal comes into existence that may alter the dynamics of the upstream system. However, by inserting an insulating device between the two, retroactivity to the upstream system is abolished and the insulating device is tuned to minimize the impact of retroactivity on its own internal dynamics.

$$\begin{aligned} \dot{x}_2 &= f_2(x_2, u_2) = f_2(x_2, y_1) \\ y_2 &= Y_2(x_2, u_2) = Y_2(x_2, y_1) \\ r &= R(x_2, u_2) \end{aligned} \tag{5}$$

Notice the former retroactivity of the input u_2, referred to as r acts as a retroactivity to the output of the insulator module S_I. At the same time, notice that the system dynamics of S_I are such that the upstream system S_1 is insulated from the effects of r; that is, no retroactivity signal maps from the insulator to S_1. In this way, the dynamics of S_I are structured to insulate the upstream module S_1 from the effects of the downstream module S_2.

The key to attenuating the retroactivity r is a design strategy of high gain coupled with negative feedback. This principle is borrowed from the design of electronic amplifiers, where retroactivity is made negligible by a theoretically infinite amplification gain and equally large negative feedback gain. This principle can be motivated using a simple linear systems model. Consider the output function Y_I of the insulator S_I, and suppose that r enters as an additive disturbance in Y_I. We suppose that negative feedback is implemented on the output y_I with gain K and that G is a transfer function describing the insulator S_I. Then supposing we can write the closed loop (insulator) dynamics of y_I in the Laplace domain, we have the following expression from [10]:

$$\begin{aligned} \hat{y}_I(s) &= G(s)\left(\hat{y}_1(s) - K(s)\hat{y}_I(s)\right) + \hat{r}(s) \\ \hat{y}_I(s) &= G(s)\left(\hat{u}_I(s) - K(s)\hat{y}_I(s)\right) + \hat{r}(s) \end{aligned} \tag{6}$$

which can be written as

$$\hat{y}_I(s) = \frac{G(s)}{1 + K(s)G(s)}\hat{u}_I(s) + \frac{1}{1 + K(s)G(s)}\hat{r}(s). \tag{7}$$

By increasing the gain of either G or negative feedback gain K, we can render the contribution from the retroactivity r negligible. Finally, as we increase the system gain of S_I, namely the gain of G, the signal $\hat{y}_I(s)$ tends towards $\hat{u}(s)/K$. So far, these observation only implicate high-level design specifications on the phosphorylation based insulator. In [10], these design specifications are shown to be satisfied by two types of insulators, a

transcriptional feedback insulator and a phosphorylation based insulator. Our paper focuses on modeling and characterization of the latter in the TXTL system. We now briefly discuss the experimental implementation of the phosphorylation based insulator in the TXTL system; a more detailed treatment is given in [15].

Based on a design postulated in [10], we constructed an adaptation of the phosphorylation based insulator to implement in the TXTL system. The insulator design is based on a well-known two component signal transduction system regulating the transcription of genes encoding metabolic enzymes and permeases. These genes are activated (or repressed) in response to carbon and nitrogen in *Escherichia coli* and related bacteria [16]. There are two essential proteins in the system, NRII and NRI (NtrB-NtrC). NRI is phosphorylated into NRI^P by $NRII$ (which assumes the conformational state of a kinase). Only NRI^P is able to activate the σ^{54}-dependent promoter p_{glnA} and trigger the transcription of downstream genes [17]. NRII is both a kinase and a phosphatase, regulated by the PII signal transduction protein, which, on binding to NRII, inhibits the kinase activity of NRII and activates the NRII phosphatase activity [18]. NRII is known to bind to itself to form a dimer as well as participate in autophosphorylation to become a kinase. Previous studies suggested that when NRII has a mutation of leucine to arginine at residue 16, it loses its phosphatase activity but showed normal autophosphorylation. In contrast, NRII has a H139N mutation, which renders it unable to autophosphorylate itself. Thus, NRIIL16R only acts as a kinase and NRIIH139N only functions as a phosphatase [19].

In our TXTL implementation of the phosphorylation based insulator, we ensure that NRI, NRIIL16R, and NRIIH139N are expressed constitutively with p_{Lac} promoter in the absence of LacI protein. A reporter molecule, deGFP [11], [12] is controlled by a $\sigma - 54$ dependent promoter which we will refer to as p_{GlnA}. The p_{GlnA} promoter is activated by phosphorylated NRI, or NRI^P. Finally, the NRIIL16R acts as a kinase to phosphorylate NRI and NRIIH139N acts as a phosphatase to dephosphorylate NRI^P.

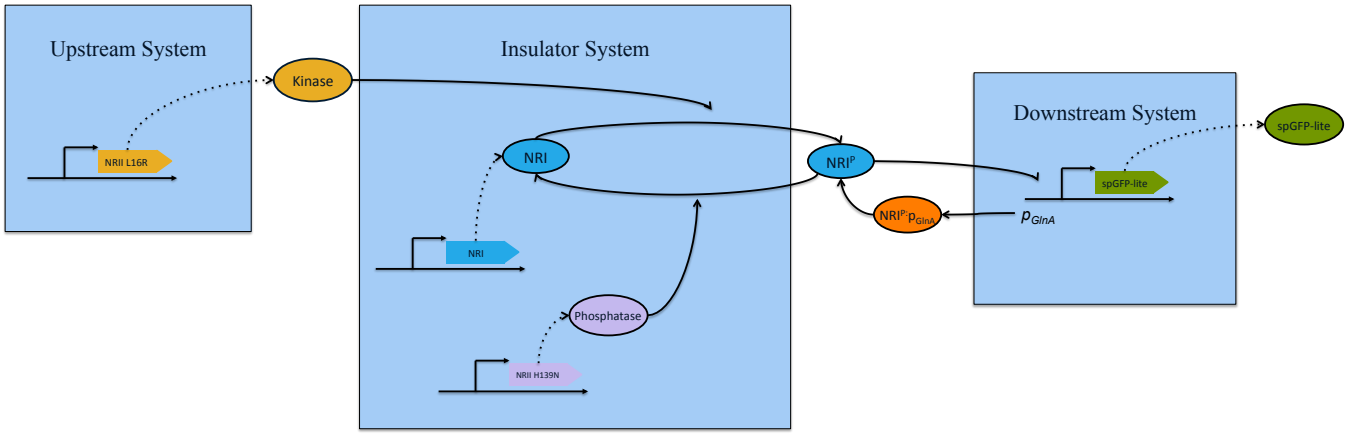

Fig. 2: A schematic illustrating the gene regulatory network in the phosphorylation-based insulator.

III. Estimation of Constitutively Expressed Protein Concentrations

In this section, our goal is to derive a simplified model that can be parameterized uniquely from a set of characterization experiments in the bimolecular breadboard system [11], [12]. We base our model on the general phosphorylation based insulator model posed in [5], but adapt our notation and augment input variables that are present in the biomolecular breadboard system. Because it is an *in vitro* system, the total DNA and inducer concentration in solution are adjustable experimental variables or variables that can be modeled as inputs. It is the freedom of these inputs that allows us to perform experiments and collect data that parameterizes the model.

We begin by introducing a chemical reaction model for the system:

$$
\begin{aligned}
NRI + K &\xrightarrow{k_{ph}} NRI^P + K \\
NRI^P + Ph &\xrightarrow{k_{deph}} NRI + Ph \\
NRI^P &\xrightarrow{k_{auto}} NRI \\
p_K &\xrightarrow{k_{TX,K}} m_K \xrightarrow{k_{TL,K}} K^u \xrightarrow{k_{f,K}} K \\
p_{Ph} &\xrightarrow{k_{TX,Ph}} m_{Ph} \xrightarrow{k_{TL,Ph}} Ph^u \xrightarrow{k_{f,Ph}} Ph \\
p_N &\xrightarrow{k_{TX,N}} m_{NRI} \xrightarrow{k_{TL,N}} NRI^u \xrightarrow{k_{f,N}} NRI \\
m_{NRI} &\xrightarrow{\delta_m} \emptyset \\
m_K &\xrightarrow{\delta_m} \emptyset \\
m_P &\xrightarrow{\delta_m} \emptyset \\
NRI^P + p_{GlnA} &\underset{k_u}{\overset{k_b}{\leftrightarrows}} NRI^P : p_{GlnA} \xrightarrow{k_{cat}} NRI^P + p_{GlnA} + deGFP \\
NRI^P + p_{GlnA,L} &\underset{k_u}{\overset{k_b}{\leftrightarrows}} NRI^p : p_{GlnALoad} \xrightarrow{k_{cat}} NRI^P + p_{GlnA,L} + RFP
\end{aligned}
\tag{8}
$$

where K, P, NRI, NRI^P denote the kinase, phosphatase, unphosphorylated NRI and phosphorylated NRI protein, p_{GlnA} is the GlnA promoter, $p_{GlnA,L}$ is the GlnA promoter encoding for other competing genes, NRI^u is the unfolded form of NRI protein, K^u is the unfolded form of kinase, Ph^u is the unfolded form of dephosphatase, and $\emptyset$ represents a macrostate of all degraded mRNA. We also use the notation X^{tot} when needed to denote the total amount of protein X where $X = NRI, K,$ and Ph. This notation will be convenient for our analysis in the sequel.

Since p_X expresses as a constitutive promoter for $X = K, Ph, N$ (short for NRI), the total kinase, phosphatase, and NRI protein are produced constitutively. An assay with green fluorescent protein (deGFP) shows that without additional proteases added into the bimolecular breadboard system, protein degradation is near negligible (see the right subfigure in Figure 4). Thus, we can approximate the total amount of NRI protein at a particular time t expressed under the p_{Lac} promoter using deGFP expression expressed under the p_{Lac} as a proxy. This total amount of NRI, we will denote as NRI^{tot}.

In taking this approach, we wish to clarify that the corresponding assay uses a deGFP reporter molecule expressed on an isolated linear form of DNA, distinct from the DNA encoding the p_{GlnA} promoter and deGFP coding sequence used in the phosphorylation based insulator. Thus, in what follows, our reference to deGFP will refer to the protein expressed on linear DNA, as a single isolated gene in the transcription-translation system.

We also note that an alternative approach to estimate $NRI^{tot}(t)$ is to assay the expression of a NRI-GFP fusion protein. However this approach may significantly alter the phosphorylation dynamics of the NRI protein, since it acts as a substrate for the kinase. Therefore, we will express deGFP separately on the p_{Lac} promoter and use a calibration curve to estimate concentration from arbitrary units of fluorescence.

Because there are differences in the folding, transcription, and translation rates of deGFP and NRI, we do not expect the estimated concentration of GFP at time t will be identical to the concentration of the NRI protein at time t. We can account for these differences dynamically in an mass action model of NRI and deGFP dynamics. If we consider NRI constitutive expression in a simple isolated system with no kinase or phosphatase activity

NRI^{tot}, e.g with the chemical reaction system

$$p_N \xrightarrow{k_{TX,N}} m_{NRI} \xrightarrow{k_{TL,N}} NRI^u \xrightarrow{k_{f,N}} NRI \tag{9}$$

we see that

$$\begin{aligned} \dot{NRI} &= k_{f,N} NRI^u \\ \dot{NRI^u} &= k_{TL,N} m_{NRI} \\ \dot{m}_{NRI} &= k_{TX,N} p_{Lac} - \delta_m m_{NRI} \end{aligned} \tag{10}$$

The total NRI protein at time t is ultimately a function of $m_{NRI}(t)$. Since the dynamics of m_{NRI} can be viewed as a scalar linear system with static step input p_{Lac}, we can solve analytically for $NRI^{tot}(t)$ to obtain:

$$\begin{aligned} NRI(t) &= NRI(t_0) + \int_0^t \left[k_{f,N} NRI^u(t_0) + k_{TL,N} \int_0^\tau e^{-\delta_m \xi} m_{NRI}(t_0) + \frac{k_{TX} p_{Lac}}{\delta_m} \left(1 - e^{-\delta_m \xi}\right) \right] d\tau dt \\ &= \frac{k_{f,N} k_{TX,N} k_{TL,N}}{\delta_m} p_{Lac} \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{-\delta_m t} \right) \end{aligned} \tag{11}$$

whenever $NRI(t_0) = m_{NRI}(t_0) = NRI^u(t_0) = 0$. To reflect the experimental conditions of our system, we have assumed that the initial mRNA, unfolded and folded kinase, phosphatase and NRI concentrations are zero. Notice that in deriving this expression, we have made no assumption about time-scale separation. While such arguments are valid since the folding dynamics proceed at a much slower rate than the transcription and translation dynamics, they are unnecessary for estimating NRI at time t. Finally, it is worth noting that we assume the transcription and translation reactions proceed as first order reactions, which is valid as long as our DNA concentrations (typically in the nM range) are much less than the concentrations of RNA polymerases, ribosomes, chaperone proteins, etc. (typically in the μM range).

It is worth noting that model for the mRNA species m_{NRI} is qualiitatively consistent with our experimental studies of mSpinach expression in the transcription-translation system. To demonstrate this, we consider a model of the same functional form as (11), but with a constitutive promoter and coding sequence of the same length as the mSpinach transcript.

$$m_S(t) = \frac{k_{TX} p_{Lac}}{\delta_m} (1 - e^{-\delta_m t}) \tag{12}$$

where

$$k_{TX} = k_{r,bp} / L(mS) \times k_{isom}$$

is estimated with $k_{r,bp} = 60$ bp s^{-1} (the approximate mean of a variety of media-dependent rates found in [20]), $L(mS) = 98$ bp mSpinach^{-1} is the length of mSpinach aptamer without a tRNA scaffold [21], and $k_{isom} = 6.3 \times 10^{-2}$ s^{-1} is the forward rate of open complex formation from the closed complex.

From this, it is possible to estimate the rate of mRNA degradation,

$$\delta_m = \frac{k_{TX} p_R}{mS(t = 120)} = 3 \times 10^{-2} \text{ s}^{-1},$$

where $mS(t = 120)$ is the expression of mSpinach at time $t = 120$ and an approximation of mS steady state expression, if the system were to continue to run indefinitely. The time point $t = 120$ minutes or 2 hours, is critical to consider for our biomolecular breadboard system. After 120 minutes, the transcription translation system begins to lose functionality, including functionality of its transcription, which enables degradation to

become the predominating force in determining mRNA concentration. We see a subsequent decrease in mRNA concentration accordingly. Thus, an empirical upper bound on time horizon for our model is approximately 120 minutes after the reaction is initiated.

It is also important to mention that with the exception of the mRNA species m_{NRI} of NRI protein, in our model, the species associated with NRI do not settle at a stable steady state. This aspect of our model is consistent with the behavior of biocircuit expression for an initial window of time in the biomolecular breadboard system. In this *in vitro* system, the auxiliary proteins NRI, K, Ph and even $deGFP$ (expressed by p_{GlnA}) do not achieve a steady state in the traditional manner (due to detailed balance of production and a combination of degradation and dilution effects). Rather, they continue to increase in concentration until all transcriptional and translation resources are exhausted. Thus, because we are interested in the *dynamic* behavior of the phosphorylation based insulator and drawing comparisons of its *in vitro* behavior to *in vivo* behavior, we will focus our subsequent modeling efforts in the time window $t \in (0, 120]$ minutes where fuel, energy and other transcriptional and translational resources are still abundant. In doing so, we do not preclude the possibility of genes competing with each other for the finite resources available in the *in vitro* system. Our time frame of interest is thus when transcriptional and translational machinery is available and functional, but in finite supply (mimicking *in vivo* conditions).

Using the parameters we have calculated, we plot the outcome of a simulation against expression data for $m_{Spinach}$ in Figure 3. The output of the simulation is simulated with additive white noise, replicating the measurement noise present in the plate reader (refer to the trajectory of the negative control). We use a biocircuit expressing mSpinach with the constitutive promoter (pOR1-OR2 from the λ regulatory operon). Notice that the functional form of $m_{NRI}(t)$ adequately describes the qualitative behavior of mRNA expression in the breadboard system until $t \approxeq 120$ minutes. The rate at which mSpinach saturates is determined by the δ_m parameter and its steady state value is given as $k_{TX}p_{Lac}/\delta_m$. These experiments with the mSpinach RNA aptamer show that our model, while simple in its formulation, is sufficiently complex to describe transcriptional dynamics in the transcription-translation system for the first two hours. Thus, we will not attempt to model system expression when the transcription-translation system depletes it resources; at this point gene expression is strongly competitive, production and degradation rates are largely determined by the available ATP, rNTP, amino acids, etc. in the system.

We also know that the folding, transcription, and translation rates of the deGFP protein are different from those of NRI protein. Thus, to estimate the quantity $NRI^{tot}(t)$ from $deGFP(t)$ To estimate the ratio in folding rates, we use the K-fold protein folding simulation software developed in [22]. We also assume that the primary variables that determines change in transcription or translation rates are length of coding sequence and length of reading frame, respectively. We express the rates of transcription, translation, and folding for NRI in terms of GFP rates of transcription, translation, and folding (respectively) as follows:

$$k_{TX,N} = \frac{725\,_{bp/GFP}}{1457\,_{bp/NRI}} k_{TX,G} \equiv \alpha_{TX} k_{TX,G}$$
$$k_{TL,N} = \frac{226\,_{res/mGFP}}{470\,_{res/NRI}} k_{TL,G} \equiv \alpha_{TL} k_{TL,G}$$
$$k_{f,N} = \frac{.85\,s^{-1}}{1.23\,s^{-1}} k_{f,G} \equiv \alpha_f k_{f,G}$$

where '$\equiv$' denotes a definition of α_j. Our model for GFP expression under the p_{Lac} promoter is similarly

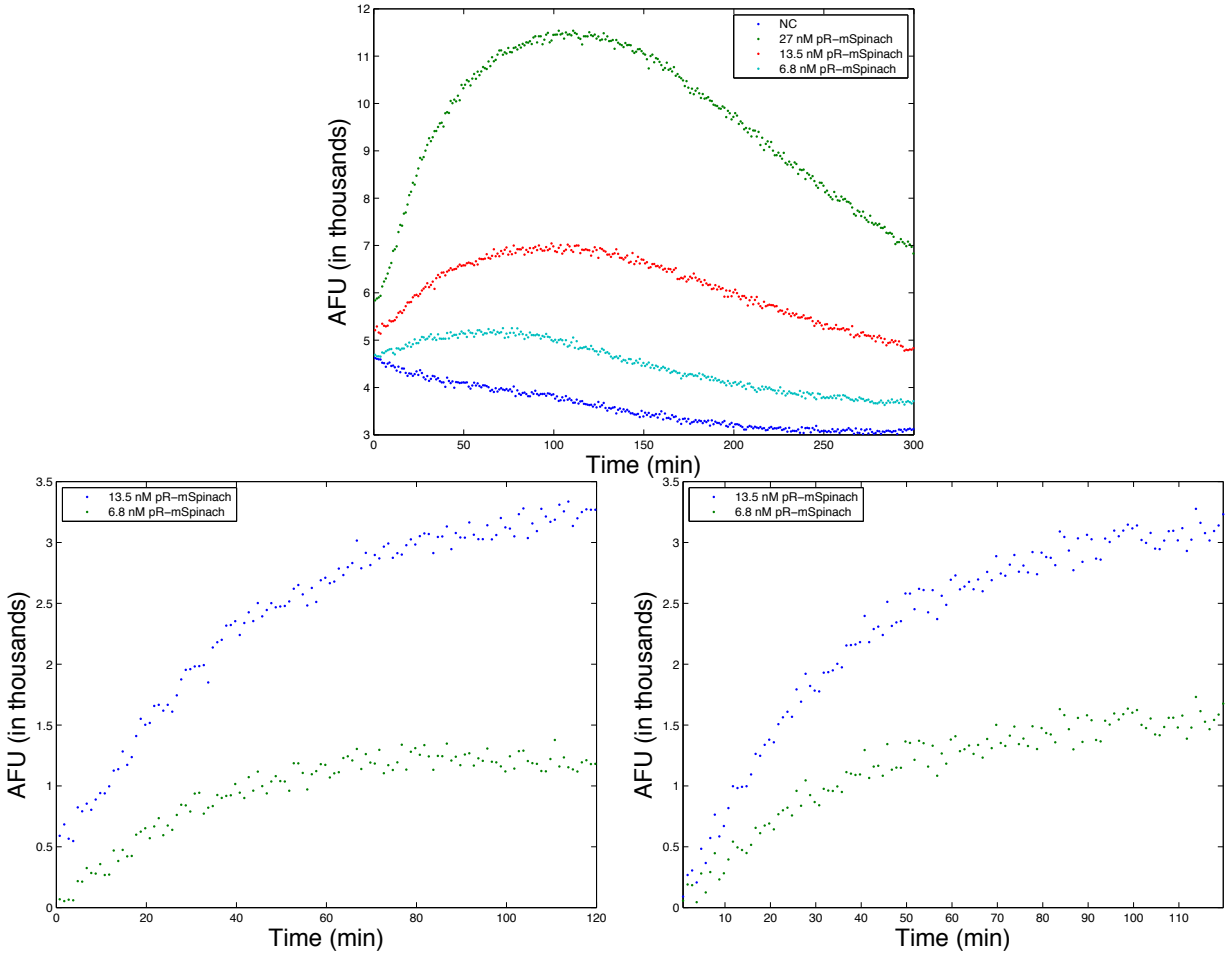

Fig. 3: (Top) Data featuring mSpinach expression on linear DNA with 100 bp of protection. The transcriptional unit consists of an OR1-OR2-pR promoter, followed by the mSpinach (no scaffold) RNA aptamer coding sequence [21] and the T500 terminator [12]. Arbitrary fluorescence units of mSpinach expression is plotted against time. Subtracting the background, we see that mSpinach expression nearly doubles as DNA concentration doubles. Past $t = 120$ minutes, mSpinach expression decreases, presumably because linear DNA template has degraded or transcriptional resources are exhausted. Our time horizon of interest for the model will thus be in the interval of $t \in (0, 120]$ (Left) Data featuring mSpinach expression driven by the OR1-OR2-pR promoter at 13.5 nM and 6.8 nM concentration from the time interval of 0 to 120 minutes. mSpinach expression dynamics in the time horizon of interest feature a phase of steep linear growth and then saturation towards an asymptotic limit. (Right) A simulation of mSpinach expression, driven by a constitutive promoter at 6.8 and 13.5 nM DNA concentration, based on the model (12). Notice that the model is able to capture the qualitative effects of mSpinach expression.

expressed as:

$$p_{Lac} \xrightarrow{k_{TX,G}} m_{GFP} \xrightarrow{k_{TL,G}} GFP^{u} \xrightarrow{k_{f,G}} GFP^{tot}$$

The model derived for $GFP(t)$ follows an analogous derivation as the model for $NRI^{tot}(t)$. Thus using equation (11), it is straightforward to show that $NRI^{tot}(t)$ concentration can be expressed as

$$\begin{aligned} NRI^{tot}(t) &= \frac{\alpha_{TX} k_{TX,G} \alpha_{TL} k_{TL,G} \alpha_f k_{f,G} p_{Lac}}{\delta_m} \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{\delta_m t} \right) \\ &= \alpha_{TX} \alpha_{TL} \alpha_f GFP(t) \end{aligned} \tag{13}$$

We see that by scaling the GFP concentration by the appropriate ratios at time t, we can obtain an estimate for NRI^{tot}. The above formula holds as long as the concentration of p_{Lac} promoter expressing deGFP is the same as the concentration of p_{Lac} promoter expressing NRI protein. Otherwise, a ratio to account for the scaling between the two should also be incorporated into the above relation.

To summarize, we have posed a basic model for constitutive expression of NRI protein; the model has a closed form analytical expression that allows estimation of total NRI protein as a function of time. Our model relies on a basic set of chemical reactions describing the processes of transcription and translation. To justify our model at the transcriptional level, we have performed an experimental assay using the mSpinach RNA aptamer to ascertain the dynamics of mRNA expression in the biomolecular breadboard system. Our simulations and experimental data appear to match for up to the first two hours of the experiment, based on parameters extracted from various references, suggesting that our model is accurate in a time horizon of interest. We thus restrict our attention to this time horizon, as it represents the horizon in which transcription and mRNA degradation proceed unperturbed. Further, evidence in [23], [24] suggests that ribosomal activity proceeds unhindered in the first two hours.

We also observe that an analogous line of reasoning can be applied to estimating Ph^{tot} and K^{tot}. We do not repeat the derivation here, as it only requires a change in notation. However we emphasize that because of these observations, in the sequel we will refer to $Ph^{tot}(t)$, $K^{tot}(t)$, and $NRI^{tot}(t)$ as additional input variables (so long as we are modeling the appropriate time horizon). Additionally, it is the ratio of Ph^{tot} and K^{tot} that matter as a functional input in the system identification process and not the individual concentrations that matter. Further, it is by levering the inputs Ph^{tot}/K^{tot}, and NRI^{tot} we are able to identify the parameters of a simplified model uniquely.

IV. Derivation of a Simplified Model for the Phosphorylation Based Insulator

In this section we derive a simplified model of the phoshorylation based insulator using the chemical reaction system (8). Examining the full chemical reaction system (8), we obtain the following state space model from the

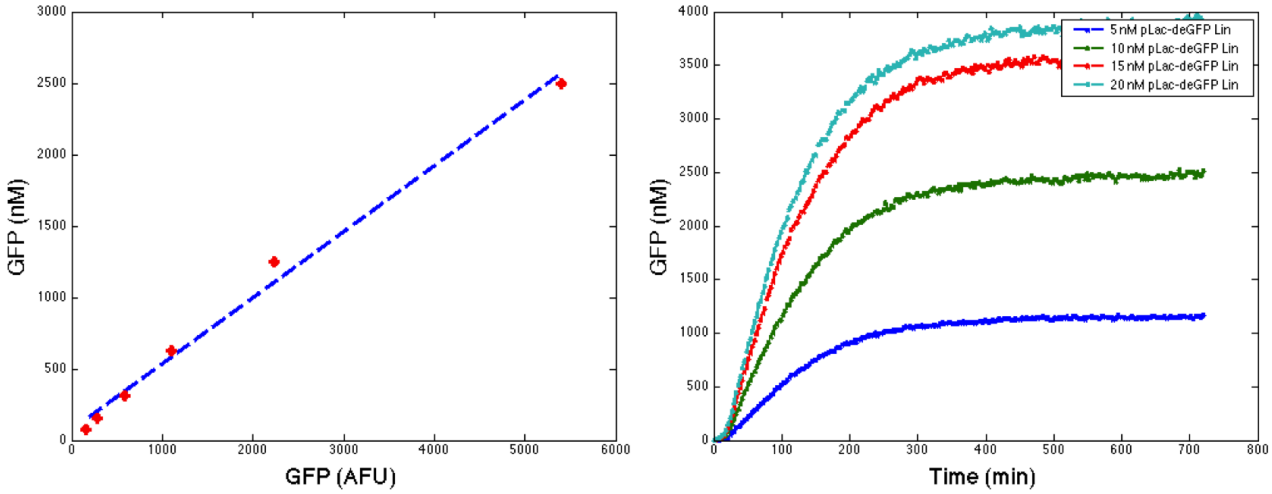

Fig. 4: (Left) A calibration curve of purified deGFP concentration against arbitrary fluorescence units of a BioteK Synergy H1 Hybrid Multi-mode Microplate Reader at 485 nm excitation wavelength, 525 nm emission wavelength, measured at 29° C with a gain parameter of 61. deGFP protein was labeled with a His tag and purified using protein affinity chromatography. (Right) pLac-deGFP expression over a range of DNA concentrations in the cell free transcription-translation system. Measurements of bulk reactions at 10 μL volumes were taken in the BioteK Synergy H1 Hybrid Multi-mode Microplate Reader at 485 nM, 525 nm with gain 61 at 29° C. Increasing DNA concentration produces an increase in deGFP expression. In particular, the curve for pLac-deGFP expression at 10 nM can be used to estimate the expression of NRI^{tot} for $t \leq 120$ min.

law of mass action:

$$
\begin{aligned}
\dot{NRI^P} &= k_{ph}\, NRI\, K - k_{deph} NRI^P Ph - k_b NRI^P (p_{GlnA} + p_{GlnALoad}) \\
&\quad + (k_u + k_{cat})(NRI^p : p_{GlnA} + NRI^p : p_{GlnALoad}) \\
\dot{NRI} &= k_{deph} NRI^P Ph - k_{ph}\, NRIK + k_{f,N} NRI^u \\
\dot{NRI^u} &= k_{TL,N} m_{NRI} \\
\dot{m}_{NRI} &= k_{TX,N} p_{NRI} - \delta_m m_{NRI} \\
\dot{K} &= k_f K^u \\
\dot{K^u} &= k_{TL,K} m_K \\
\dot{m}_K &= k_{TX,K} p_K - \delta_m m_K \\
\dot{Ph} &= k_f Ph^u \\
\dot{Ph^u} &= k_{TL,Ph} m_{Ph} \\
\dot{m}_{Ph} &= k_{TX,Ph} p_{Ph} - \delta_m m_{Ph} \\
\dot{NRI^P} : p_{GlnA} &= k_b NRI^P p_{GlnA} - (k_u + k_{cat})(NRI^P : p_{GlnA}) \\
\dot{NRI^P} : p_{GlnALoad} &= k_b NRI^P p_{GlnALoad} - (k_u + k_{cat})(NRI^P : p_{GlnALoad}) \\
\dot{GFP} &= k_{cat} NRI : p_{GlnA} \\
\dot{RFP} &= k_{cat} NRI : p_{GlnALoad}
\end{aligned}
\tag{14}
$$

The dimension of the state-space model is fourteen and because of the presence of bimolecular reactions, it

is nonlinear in the state of the system. Thus, it is difficult to obtain a closed form expression for the solution to the system. However, we will systematically impose a series of modeling assumptions that are physiologically plausible, but which greatly reduce the complexity of the model.

First, notice that the total concentration of K, Ph and NRI, denoted as K^{tot}, Ph^{tot}, depends only on the transcription and translation reactions. Thus, if we consider the transcription and translation dynamics of K, Ph and

$$NRI^{tot} = NRI^P + NRI + NRI^P : p_{GlnA} + NRI^P : p_{GlnALoad} = k_{f,N} NRI^u$$

in isolation, we can use the results of the previous section to obtain a closed form expression for their total concentration as follows:

$$\begin{aligned} NRI^{tot}(t) &= \frac{k_{f,N} k_{TX,N} k_{TL,N}}{\delta_m} p_{NRI} \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{-\delta_m t} \right) \\ K^{tot}(t) &= \frac{k_{f,N} k_{TX,K} k_{TL,K}}{\delta_m} p_K \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{-\delta_m t} \right) \\ Ph^{tot}(t) &= \frac{k_{f,Ph} k_{TX,Ph} k_{TL,Ph}}{\delta_m} p_{Ph} \left(\frac{t^2}{2} - \frac{1}{\delta_m^2} e^{-\delta_m t} \right) \end{aligned} \tag{15}$$

These total concentrations can be viewed as time varying parameters. If we had a way of quantifying the rate of transcription, translation, and folding of the individual proteins in the transcription-translation system, we could predictively estimate the trajectories of NRI^{tot}, K^{tot}, and Ph^{tot} over time. However, we do not have these parameters, and thus it is advantageous to employ the previous section's approach. With similar arguments, we can argue that the total concentration of these proteins can be expressed as the total concentration of a reporter molecule multiplied by a scaling constant (see equation (13)). Thus, using a separate assay to quantify constitutive expression of a reporter molecule under a given constitutive promoter (and a calibration curve to convert fluorescence to molar concentration), we can use the reporter molecule as a proxy for estimating the true molar concentration of NRI, kinase or phosphatase. Therefore, we can avoid the problem of estimating transcriptional, translational and folding rates of heterogeneous proteins while obtaining an estimate of the functional protein concentrations. Moreover, the result holds for all t in which RNA expression increases linearly ($t \leq 120$). We formalize this assumption as follows:

Assumption 1: We suppose that for all $t \in [0, 120]$, $K^{tot}(t), Ph^{tot}$, and $NRI^{tot}(t)$ are known parameters.

This assumption thus allows us to eliminate the dynamics of folded kinase, unfolded kinase, folded NRI, unfolded NRI, folded phosphatase, unfolded phosphatase and all mRNA dynamics.

The remaining dynamics of the system are thus given as:

$$\begin{aligned} \dot{NRIP} &= k_{ph}\, NRI\, K - k_{deph} NRI^P Ph - k_b NRI^P (p_{GlnA} + p_{GlnALoad}) \\ &\quad + (k_u + k_{cat})(NRI^p : p_{GlnA} + NRI^p : p_{GlnALoad}) \\ \dot{NRI} &= k_{deph} NRI^P Ph - k_{ph}\, (NRI)(K) + k_{f,N} NRI^u \\ \dot{NRI}^P : p_{GlnA} &= k_b NRI^P p_{GlnA} - (k_u + k_{cat})(NRI^P : p_{GlnA}) \\ \dot{NRI}^P : p_{GlnALoad} &= k_b NRI^P p_{GlnALoad} - (k_u + k_{cat})(NRI^P : p_{GlnALoad}) \\ \dot{GFP} &= k_{cat} NRI^p : p_{GlnA} \\ \dot{RFP} &= k_{cat} NRI^p : p_{GlnALoad} \end{aligned} \tag{16}$$

Next, we assume that the phosphorlyation and dephosphorylation reactions occur at a much faster time scale

then production of GFP or RFP and the binding (and unbinding) reactions of NRI^P to DNA to form (or disintegrate) activator-DNA complex. We justify the latter assumption through experimental observations that observe phosphorylation rates on the order of 10^6 min^{-1}. Transcription factor binding rates are less characterized but typically binding and unbinding rates of a transcription factor (e.g. LacI) are O(10^{-1}) min^{-1} and $O(10)$ min^{-1} respectively [25]. We formalize these assumptions as follows

Assumption 2: We suppose that

$$k_{ph}, k_{deph} \gg k_u, k_{cat}, k_b$$

Next, we suppose that the amount of DNA bound NRI^p is smaller than the amount of free NRI^p and unphosphorylated NRI and that total NRI can be approximated as the sum of unbound NRI^p and NRI. Put another way, we assume that the molar concentration of unbound NRI protein is substantially larger than the molar concentration of DNA bound NRI protein. This will certainly be the case since the p_{GlnA} and $p_{GlnAload}$ DNA concentration will be in the nanomolar range while the protein concentration of NRI will be in the micromolar range (refer to the arguments in the previous section and Figure 3). From the above reactions and assumptions, we then can write the dynamics of NRI^P using the approximate conservation law $NRI^{tot} \approxeq NRI^p + NRI$ as follows:

$$\dot{NRI^P} = k_{ph}(NRI^{tot} - NRI^P)K^{tot} - k_{deph}Ph^{tot}NRI^P$$

Since phosphorylation and dephosphorylation occurs at a much faster rate than GFP and RFP production (our ultimate time-scale of interest) and reasonably faster than the binding dynamics of the NRI^P transcriptional activator, we can solve the fast dynamics of NRI^P to obtain an analytical expression for the equilibrium point NRI_e^P. At steady state, we have

$$0 = \dot{NRI^P} = k_{ph}(NRI^{tot} - NRI^P)K^{tot} - k_{deph}Ph^{tot}NRI^P$$

which implies

$$\begin{aligned} NRI_e^P &= \frac{k_{ph}NRI^{tot}K^{tot}}{k_{ph}K^{tot} + k_{deph}Ph^{tot}} \\ &= \frac{k_{ph}/k_{deph}NRI^{tot}}{k_{ph}/k_{deph} + (Ph^{tot}/K^{tot})} \\ &\equiv \theta(NRI^{tot}, K^{tot}, Ph^{tot}). \end{aligned}$$

where '$\equiv$' denotes the definition of the function $\theta(NRI^{tot}, Ph^{tot})$. The careful reader comparing the model of [5] and our simplified model will observe that we have assumed k_{auto} is negligible. This is a reasonable assumption since spontaneous dephosphorylation proceeds at a slow rate — the Δ_G of spontaneous dephosphorylation is very large [19]. The final assumption we leverage is that the rates of GFP and RPF production, relative to the binding dynamics of NRI^P are much slower. Specifically, we suppose that:

Assumption 3:

$$k_b, k_u \gg k_{cat}$$

This assumption can be justified, since the production of a folded protein such as GFP takes at least ten to fifteen minutes [26] while the binding and unbinding rates are typically on the order of hundreds of seconds and seconds, respectively [25]. Thus, we can solve for the steady state of the DNA-activator complexes $NRI^P : p_{GlnA}$ and $NRI^P : p_{GlnALoad}$. The result is analogous to the classical Michaelis-Menten model, with $V_{max} = NRI_e^P$ and $K_M = (k_u + k_{cat})/k_b$. We omit the derivation, as it follows the standard derivation for a two-substrate one

enzyme model:

$$\begin{aligned} de\dot{G}FP &= k_{cat}\frac{\theta p^{tot}_{\text{GlnA}}/K_M}{1+\left(p^{tot}_{\text{GlnA}}+p^{tot}_{\text{GlnALoad}}\right)/K_M} \\ R\dot{F}P &= k_{cat}\frac{\theta p^{tot}_{\text{GlnALoad}}/K_M}{1+\left(p^{tot}_{\text{GlnA}}+p^{tot}_{\text{GlnALoad}}\right)/K_M} \end{aligned} \tag{17}$$

where θ denotes $\theta(NRI^{tot}, Ph^{tot})$. This completes the derivation of our simplified model. In the next section, we will explore the analyze the structure of the model, determine which of the parameters are globally identifiable, and under what circumstances identifiability holds.

V. System Identification of the Simplified Phosphorylation Based Insulator Model

A. Theoretical Analysis

In the derivation of our model we have made a point to retain the experimental parameters $NRI^{tot}, K^{tot}, Ph^{tot}$, p^{tot}_{GlnA}, and $p^{tot}_{\text{GlnALoad}}$. These parameters can be viewed as experimentally controllable, in that we can directly control the DNA concentration of promoters p^{tot}_{GlnA} and $p^{tot}_{\text{GlnALoad}}$. Additionally, by adjusting the underlying constitutive promoters driving the expression of NRI, K, and Ph we can effectively tune the quantities NRI^{tot}, K^{tot}, and Ph^{tot}. We note this type of control over the concentration of DNA as well as total protein concentrations is not typically achievable *in vivo*, unless inducers are employed (which introduce stochastic effects from membrane diffusion) or different replication origins are cloned into a plasmid (which introduces variability in copy number from cell-to-cell). However, this advantage in the biomolecular breadboard is precisely the capability required to explore the problem of parameter estimation and determine if our simplified model is globally identifiable.

Since our calibration curves allow us to estimate GFP concentration from arbitrary fluorescence units, we will focus our attention on the GFP dynamics. Furthermore, notice that the dynamics of both reporter molecules are identical. Thus, it suffices to analyze the identifiability of parameters with respect to the output dynamics of the GFP reporter molecule, since it will yield the same result as studying identifiability with respect to RFP output dynamics. Recalling our assumptions from the previous section, we will also make a point to study the behavior of the system within the time horizon of interest captured by our model, $t \in [0, \tau_{max}]$. where τ_{max} is the initiation of the resource depletion phase in our transcription-translation system.

Our goal is to determine whether this model is globally structurally identifiable with respect to the parameters k_M, k_{cat}, k_{ph}, and k_{deph}, given the inputs $NRI^{tot}, K^{tot}, Ph^{tot}, p^{tot}_{\text{GlnA}}$, and $p^{tot}_{\text{GlnALoad}}$. Notice that the inputs $NRI^{tot}, Ph^{tot}, K^{tot}, p^{tot}_{GlnA}$, and $p^{tot}_{GlnALoad}$ do not enter the dynamics of the system in a linear fashion. Indeed, the simplified system 17 is of the form:

$$\dot{x} = f(U, \Theta) \tag{18}$$

where f is nonlinear with respect to U and Θ and

$$U = \left(NRI^{tot}, Ph^{tot}, K^{tot}, p^{tot}_{GlnA}, p^{tot}_{GlnALoad}\right)$$

and

$$\Theta = \left(k_M, k_{cat}, k_{ph}, k_{deph}\right).$$

Furthermore,

$$f(U, \Theta) = f_1(U_1, \Theta_1) f_2(U_2, \Theta_2)$$

where $U_1 = (NRI^{tot}, Ph^{tot}, K^{tot}), \Theta_1 = (k_{ph}, k_{deph}), U_2 = (p^{tot}_{GlnA}, p^{tot}_{GlnALoad})$ and $\Theta_2 = (k_{cat}, K_M)$. Notice that $f_1 = NRI^P_e = \theta(NRI^{tot}, K^{tot}, Ph^{tot})$ takes the form of a Hill function with Ph^{tot}/K^{tot} as its substrate (i.e. argument) and

$$f_2 = k_{cat} \frac{p^{tot}_{\text{GlnA}}/K_M}{1 + \left(p^{tot}_{\text{GlnA}} + p^{tot}_{\text{GlnALoad}}\right)/K_M}.$$

This multiplicative decomposition provides a key insight: our system dynamics is the product of two Hill functions with distinct inputs for each Hill function. This suggests that from a system identification standpoint, we can attempt a series of experiments that perturb one of the Hill functions while holding the other constant and vice versa to tease out the parameters for each.

To obtain insight into the what parameters in the Hill functions are identifiable, we invert the system dynamics, we obtain

$$\begin{aligned} \frac{1}{G\dot{F}P} &= \frac{1}{f_1 f_2} \\ &= \frac{(p^{tot}_{GlnA} + p^{tot}_{GlnALoad}) + K_M}{k_{cat}\theta p^{tot}_{GlnA}} \end{aligned} \tag{19}$$

and after some rearrangement, we obtain that

$$\frac{G\dot{F}P(p^{tot}_{GlnA} + p^{tot}_{GlnALoad})}{p^{tot}_{GlnA}} = -K_M f_1 \frac{G\dot{F}P}{p^{tot}_{GlnA}} - f_1 k_{cat}. \tag{20}$$

Thus, when the experimental input $p^{tot}_{GlnALoad}$ is set to 0 nM, we obtain a linear regression problem in estimating slope $K_M f_1$ and intercept $f_1 k_{cat}$. Further, if we enforce that $Ph^{tot} = 0$, then f_1 reduces to NRI^{tot}, a known input value which completes the decomposition. Thus, by enforcing these two input constraints, we obtain a linear regression problem that effectively estimates K_M and k_{cat}. By varying the total DNA concentration p^{tot}_{GlnA} we can thus vary the rate of change of GFP, $G\dot{F}P$, and obtain data to optimize K_M and k_{cat}. Once k_{cat} and K_M are estimated, we can then use a similar line of arguments to back out an estimate for the ratio k_{ph}/k_{depth}.

In particular, we consider a nominal operating concentration of $p^{tot}_{GlnA}, p^{tot}_{GlnALoad}$ and write $\gamma = 1/f_2(U_2, \Theta_2)$ and $k_r = k_{ph}/k_{depth}$ then taking the reciprocal of $G\dot{F}P$ we obtain

$$\frac{1}{G\dot{F}P} = \frac{1}{\gamma}\left(\frac{k_r + Ph^{tot}/K^{tot}}{k_r NRI^{tot}}\right) \tag{21}$$

and defining $\tilde{Y} = \frac{\gamma NRI^{tot}}{G\dot{F}P}$ we see that

$$\tilde{Y} = 1 + \frac{1}{k_r}\frac{Ph^{tot}}{K^{tot}}. \tag{22}$$

Therefore, by transforming the problem into the reciprocal space, we see that $k_r = k_{ph}/k_{depth}$ is a uniquely identifiable parameter from the derivative of $y_2 = GFP$. That is, the problem of estimating $k_r$can be expressed as a linear regression problem with k_r as the reciprocal of the slope and an intercept of unity. The fact that we can write the parameter estimation problem for $(k_{cat}, K_M, k_{ph}/k_{depth})$ as a solution to a system of linear equations thus shows that the model is globally structurally identifiable with respect to $(k_{cat}, K_M, k_{ph}/k_{depth})$ [27].

In summary, we have derived a simplified model for the phosphorylation based insulator and shown it is globally identifiable with respect to the output trajectory of GFP. We have shown that in the theoretical scenario where a continuous trajectory of GFP can be obtained to estimate its derivative $G\dot{F}P$, the parameters k_{cat}, K_M, and $k_r = k_{ph}/k_{depth}$ can be estimated. These parameters are only estimated through a series of carefully designed

experiments in which specific TXTL controllable experimental variables are tuned. In the next section, we discuss the results of these experiments and numerical estimation of this data from time-series data.

B. Experimental Analysis: Systematic Perturbations of the Phosphorylation Based Insulator for System Identification

To identify the parameters k_{cat}, K_M, and k_r we needed to perturb the phosphorylation based insulator with the experimental variables designated in our model. In particular, we first needed to perturb the amount of p_{GlnA} promoter producing GFP in the absence of phosphatase Ph^{tot} or NRIIH139N protein. Varying the amount of p_{GlnA} promoter in the system in the absence of phosphatase would enable the estimation of k_{cat} and K_M. Intuitively, k_{cat} and K_M characterize the enzyme-substrate relationship that the activator protein NRI^P has with the p_{GlnA} promoter — coincidentally, to reveal these parameters we need to eliminate any negative feedback imposed on the activator protein by NRIIH139N phosphatase and vary the substrate concentration p^{tot}_{GlnA} to reveal the kinetic parameters.

Accordingly, we ran a set of TXTL reactions in which the DNA concentration of p_{Lac} promoter driving NRIIH139N expression was 0 nM. We varied the concentration of p_{GlnA} promoter from 0 to $57nM$, expressed on plasmid. From the time series data we extracted the first thirty minutes of expression dynamics — this time horizon constituted the time frame when amino acids, CoA, NADH, ATP, etc. were far away from the stage of complete depletion in the TXTL system. In this time horizon of interest, the expression of GFP is linear with respect to time; therefore the derivative of GFP is constant and can be fitted using the slope of a linear regression. The results of our linear regression are plotted against the time series data of the experiment in Figure 5. The estimates of the $G\dot{F}P$ in the time horizon of interest at varying concentrations of p_{GlnA} were used to fit the Hill function parameters k_{cat} and K_M, see 6.

$$\begin{aligned} k_{cat} &= 9.74 \times 10^{-2}\, min^{-1} \\ K_M &= 1.58\, nM \end{aligned} \tag{23}$$

We emphasize that the key to estimating k_{cat} and K_M is the additional freedom afforded by a control input p^{tot}_{GlnA} in perturbing the system.

The final parameter to estimate was $k_r = k_{ph}/k_{depth}$. In order to estimate k_r, we needed to fix the p_{GlnA} concentrations, i.e. the concentrations driving expression in the downstream module and, and perturb the phosphorylation based insulator. Specifically, we varied the ratio of kinase (NRIIL16R) to phosphatase (NRIIH139N) in the system by varying the ratio of DNA concentrations for the promoters driving their expression. Doing this, we obtained a series of time-lapse curves of GFP expression over a range of Ph^{tot}/K^{tot} values. Again, we extracted estimates for $G\dot{F}P$ using a linear regression over the first thirty minutes of gene expression. The resulting estimates for $G\dot{F}P$ were then plotted against varying Ph^{tot}/K^{tot} to fit a Hill function, see Figure 6. Using standard linear regression techniques, we then obtained the following estimate:

$$k_{ph}/k_{deph} = 9.81 \times 10^{-2} \tag{24}$$

The ratio k_{ph}/k_{depth} characterizes the balance of power between phosphorylation and dephosphorylation reactions — although we are unable to infer the individual parameters k_{ph} and k_{depth} we are able to conclude that dephosphorylation occurs at roughly an order of magnitude faster than phosphorylation (all other variables equal). Notice this parameter characterizes the intrinsic chemical reaction rates, rather than the flux or mass action rates

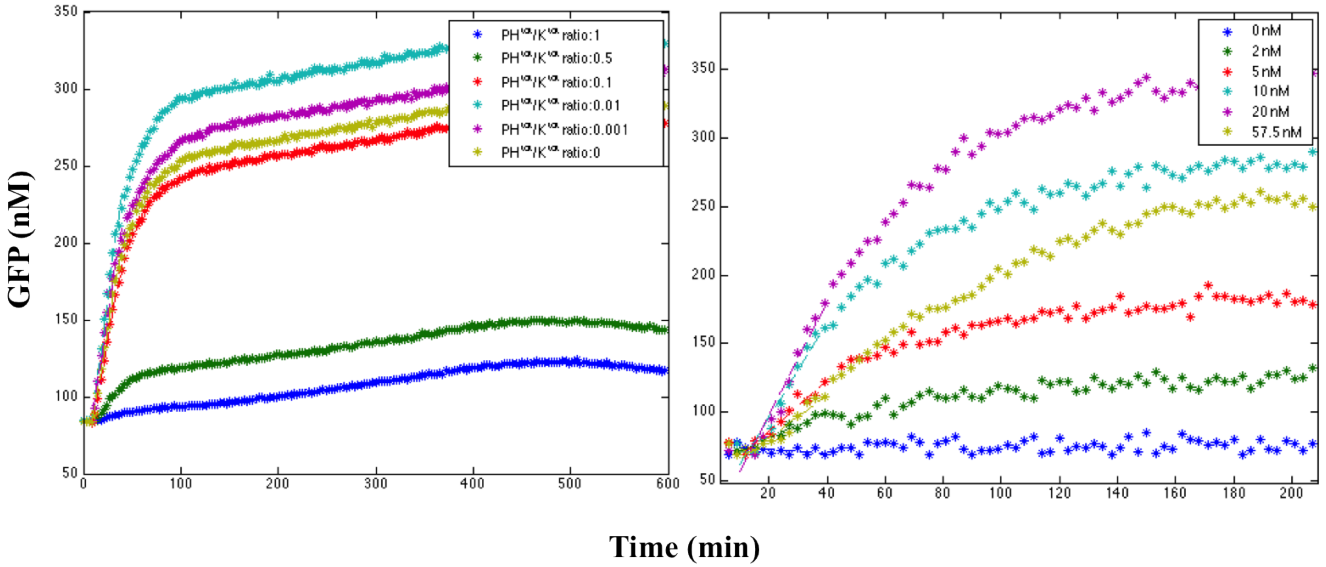

Fig. 5: (Left) Plot of GFP expression while varying the Ph over K ratio. Again, curves from $t = 0$ to $\tau_{max} = 30$ min were used to estimate the slope of GFP. (Right) Expression dynamics of GFP for varying amounts of p_{GlnA} with $pLac - Ph = 0$ nM. These curves enable the estimation of $\dot{GFP}$ for $t \leq \tau_{max} = 30min$.

that are dependent on kinase and phosphatase concentrations. Thus, to tune the phosphorylation based insulator we can vary the amount of kinase and phosphatase concentrations, bearing in mind that phosphorylation is slightly slower than dephosphorylation in the TXTL system.

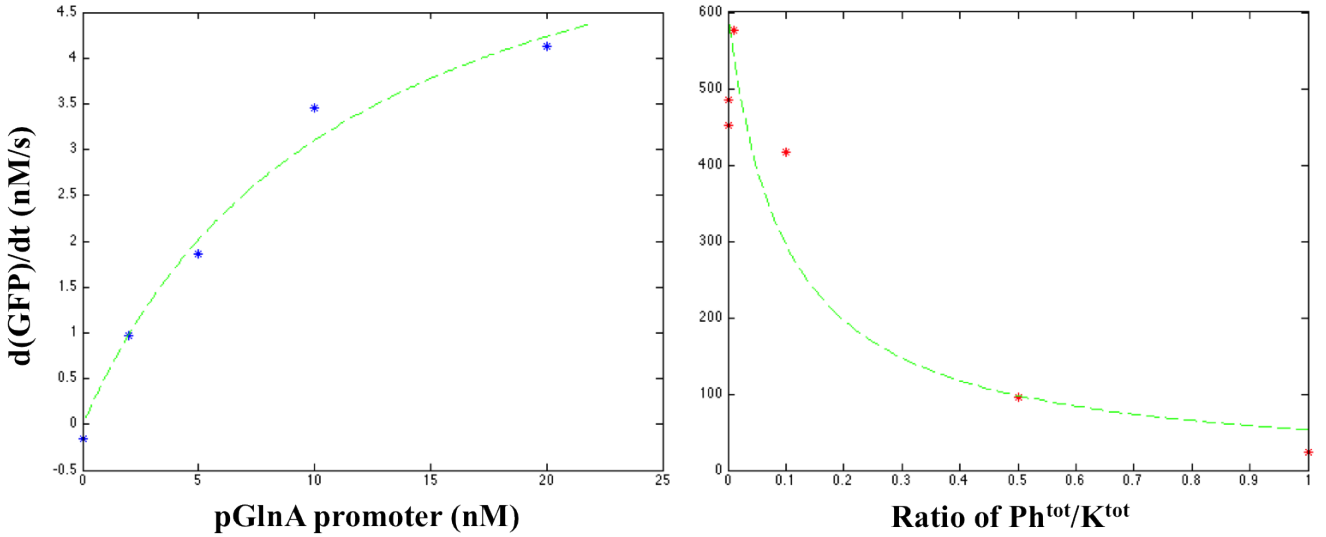

Fig. 6: (Left) A plot of the resulting Hill function $de\dot{GFP}$ against varying p_{GlnA}. We see the curve follows the form of a Michaelis Menten function which is consistent with our model. (Right) A plot of the resulting Hill function $de\dot{GFP}$ against varying Ph^{tot}/K^{tot}. Again, the empirical data (starred) matches the functional form of our model.

Further, it is consistent with our intuition that only the ratio of k_{ph} and k_{depth} is identifiable and not the individual parameters. Because the individual parameters characterize processes that are much faster than the time-scales of production of our observer molecule GFP and the imaging system in the plate reader, the only

information that can be passed onto the observer molecule is the net outcome of NRI protein's phosphorylated state. Phosphorylation and dephosphorylation are processes that compete against each other to increase the amount of NRI^p and NRI concentration in the system, respectively. Thus, by observing the amount of NRI^p in the system and knowing the concentration of NRI^{tot}, we can deduce the net outcome of the battle, i.e. the ratio k_{ph}/k_{depth}. Notice that without knowledge of NRI^{tot}, we would be unable to estimate k_{ph}/k_{depth}. This again illustrates the importance of having additional experimental inputs for perturbing the system. Even though there is only one output molecule GFP, we are able to infer three distinct parameters that represent processes from three different time-scales: catalytic synthesis of protein, formation and disassociation of the DNA-activator complex, and phosphorylation/dephosphorylation of NRI protein.

Finally, it is worthwhile to note that the functional form of our model is consistent both quantitatively (small output residual error) and quantitatively. This suggests that our simplified model will serve as a suitable starting point for simulation studies (see [15] for additional work) and theoretical analysis.

VI. Conclusion

In this work, we investigated the structural identifiability of the phoshorylation based insulator when implemented in a transcription-translation (TXTL) cell free expression system. We first considered a complex model that provided an intricate description of all chemical reactions. Next, leveraging specific physiologically plausible assumptions, we derived a rigorous simplified model that captures the output dynamics of the phosphorylation based insulator. We performed standard system identification analysis and determined that the model is globally identifiable with respect to three critical parameters: the catalytic rate associated with the downstream system k_{cat}, an internal parameter in the downstream system characterizing formation of the activator-DNA complex K_M and k_{ph}/k_{depth} a ratio describing the intrinsic balance of phosphorylation and dephosphorylation in the phosphorylation based insulator. Specifically, we showed that these three parameters were identifiable only when the system was subjected to specific perturbations. We performed these experiments and estimated the parameters. Our experimental results suggest that the functional form of our simplified model is sufficient to describe reporter dynamics and enable parameter estimation. In general, this research illustrates the utility of the TXTL cell free expression system as a platform for system identification, as it provides extra control inputs for parameter estimation that typically are unavailable in vivo. Future work will investigate the theoretical utility of the TXTL system as a platform for system identification, parameterization of more complex systems, and the robustness and sensitivity of the phosphorylation based insulator using our derived model.

References

[1] Lei Qi, Julius B Lucks, Chang C Liu, Vivek K Mutalik, and Adam P Arkin, "Engineering naturally occurring trans-acting non-coding rnas to sense molecular signals", *Nucleic acids research*, vol. 40, no. 12, pp. 5775–5786, 2012.

[2] Jerome Bonnet, Pakpoom Subsoontorn, and Drew Endy, "Rewritable digital data storage in live cells via engineered control of recombination directionality", *Proceedings of the National Academy of Sciences*, vol. 109, no. 23, pp. 8884–8889, 2012.

[3] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C Stanton, and Christopher A Voigt, "Genetic programs constructed from layered logic gates in single cells", *Nature*, vol. 491, no. 7423, pp. 249–253, 2012.

[4] J. Stricker et al, "A fast, robust, and tunable synthetic gene oscillator", *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.

[5] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag, "Modular cell biology: retroactivity and insulation", *Molecular systems biology*, vol. 4, no. 1, 2008.

[6] Peng Jiang, Alejandra C Ventura, Eduardo D Sontag, Sofia D Merajver, Alexander J Ninfa, and Domitilla Del Vecchio, "Load-induced modulation of signal transduction networks", *Science signaling*, vol. 4, no. 194, pp. ra67, 2011.

[7] Andras Gyorgy and Domitilla Del Vecchio, "Retroactivity to the input in complex gene transcription networks", in *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE, 2012, pp. 3595–3601.

[8] Hamid R Ossareh, Alejandra C Ventura, Sofia D Merajver, and Domitilla Del Vecchio, "Long signaling cascades tend to attenuate retroactivity", *Biophysical journal*, vol. 100, no. 7, pp. 1617–1626, 2011.

[9] Shridhar Jayanthi and Domitilla Del Vecchio, "Retroactivity attenuation in bio-molecular systems based on timescale separation", *Automatic Control, IEEE Transactions on*, vol. 56, no. 4, pp. 748–761, 2011.

[10] S. Jayanthi and D. Del Vecchio, "On the compromise between retroactivity attenuation and noise amplification in gene regulatory networks", *Proceedings of the 2009 IEEE Conference on Decision and Control*, December 2009.

[11] 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*, vol. 79, pp. Art–No, 2013.

[12] Jonghyeon Shin and Vincent Noireaux, "An e. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells", *ACS synthetic biology*, vol. 1, no. 1, pp. 29–41, 2012.

[13] Z. Z. Sun, E. Yeung, C. Hayes, and R. M. Murray, "Linear dna for rapid prototyping of synthetic biological circuits in an escherichia coli based tx-tl cell-free system", *preprint available at* `http://www.cds.caltech.edu/~murray/papers/2013h_sun+13-acs-synbio.html`, *to appear in ACS Synthetic Biology*, 2013.

[14] Eric Walter and Yves Lecourtier, "Unidentifiable compartmental models: What to do?", *Mathematical biosciences*, vol. 56, no. 1, pp. 1–25, 1981.

[15] S. B. Guo, E. Yeung, K. Nilgiriwala, D. Del Vecchio, and R. M. Murray, "Implementation and simulation of a phosphorylation-based insulator in a *in vitro* cell free transcription-translation system", *preprint available at* `cds.caltech.edu/~murray/papers/2013l_qkym14-wqbio.html`, *submitted to Proceedings of the 2014 Quantitative Biology Conference*, February 2014.

[16] Augen A Pioszak and Alexander J Ninfa, "Genetic and biochemical analysis of phosphatase activity of escherichia coli nrii (ntrb) and its regulation by the pii signal transduction protein", *Journal of bacteriology*, vol. 185, no. 4, pp. 1299–1315, 2003.

[17] Alexander J Ninfa, Lawrence J Reitzer, and Boris Magasanik, "Initiation of transcription at the bacterial¡ i¿ glnap2¡/i¿ promoter by purified e. coli components is facilitated by enhancers", *Cell*, vol. 50, no. 7, pp. 1039–1046, 1987.

[18] Alexander J Ninfa and Boris Magasanik, "Covalent modification of the glng product, nri, by the glnl product, nrii, regulates the transcription of the glnalg operon in escherichia coli", *Proceedings of the National Academy of Sciences*, vol. 83, no. 16, pp. 5909–5913, 1986.

[19] Augen A Pioszak and Alexander J Ninfa, "Mutations altering the n-terminal receiver domain of nri (ntrc) that prevent dephosphorylation by the nrii-pii complex in escherichia coli", *Journal of bacteriology*, vol. 186, no. 17, pp. 5730–5740, 2004.

[20] Ulla Vogel and Kaj Frank Jensen, "The rna chain elongation rate in escherichia coli depends on the growth rate.", *Journal of bacteriology*, vol. 176, no. 10, pp. 2807–2813, 1994.

[21] Jeremy S Paige, Karen Y Wu, and Samie R Jaffrey, "Rna mimics of green fluorescent protein", *Science*, vol. 333, no. 6042, pp. 642–646, 2011.

[22] Emidio Capriotti and Rita Casadio, "K-fold: a tool for the prediction of the protein folding kinetic order and rate", *Bioinformatics*, vol. 23, no. 3, pp. 385–386, 2007.

[23] D. Siegal-Gaskins, V. Noireaux, and R. M. Murray, "Biomolecular resource utilization in elementary cell-free gene circuits", *The Proceedings of the IEEE American Control Conference, to appear*, June 2013.

[24] Dan Siegal-Gaskins, Zoltan Tuza, Jongmin Kim, Vincent Noireaux, and Richard M. Murray, "Resource usage and gene circuit performance characterization in a cell-free 'breadboard"', *ACS Synthetic Biology, manuscript in preparation*, 2014.

[25] Johan Elf, Gene-Wei Li, and X. Sunney Xie, "Probing transcription factor dynamics at the single-molecule level in a living cell", *Science*, vol. 316, no. 5828, pp. 1191–1194, 2007.

[26] Jennifer A Sniegowski, Jason W Lappe, Hetal N Patel, Holly A Huffman, and Rebekka M Wachter, "Base catalysis of chromophore formation in arg96 and glu222 variants of green fluorescent protein", *Journal of Biological Chemistry*, vol. 280, no. 28, pp. 26248–26255, 2005.

[27] Eric Walter and Luc Pronzato, "Identification of parametric models", *Communications and Control Engineering*, 1997.