bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Translation Inhibition and Resource Balance in the Cell-Free Gene Expression System

Vijayalakshmi H. Nagaraj[*], James M. Greene[†], Anirvan M. Sengupta[‡], Eduardo D. Sontag[§]

May 23, 2017

Abstract

Quantifying the effect of vital resources on transcription and translation helps to understand the degree to which the concentration of each resource must be regulated for achieving homeostasis. Utilizing the synthetic transcription-translation (TX-TL) system, we study the impact of nucleotide triphosphates (NTPs) and magnesium ($Mg2^+$), on gene expression. Recent observations of the counterintuitive phenomenon of suppression of gene expression at high NTP concentrations have led to the speculation that such suppression is due to the consumption of resources by transcription, hence leaving fewer resources for translation. In this work, we investigate an alternative hypothesis: direct suppression of the translation rate via stoichiometric mismatch in necessary reagents. We observe NTP-dependent suppression even in the early phase of gene expression, contradicting the resource limitation argument. To further decouple the contributions of transcription and translation, we performed gene expression experiments with purified mRNA. Simultaneously monitoring mRNA and protein abundances allowed us to extract a time-dependent translation rate. Measuring translation rates for different $Mg2^+$ and NTP concentrations, we observe a complex resource dependence. We demonstrate that translation is the rate-limiting process that is directly inhibited by high NTP concentrations. Additional $Mg2^+$ can partially reverse this inhibition. In several experiments, we observe two maxima of the translation rate viewed as a function of both $Mg2^+$ and NTP concentration, which can be explained in terms of an NTP-independent effect on the ribosome complex and an NTP- $Mg2^+$ titration effect. The non-trivial compensatory effects of abundance of different vital resources signals the presence of complex regulatory mechanisms to achieve optimal gene expression.

Keywords: cell-free systems; in vitro synthetic biology; RNA aptamer; translation rate; homeostasis

[*]Center for Quantitative Biology (CQB), Rutgers University, Piscataway, NJ 08854, United States
[†]Department of Mathematics and CQB, Rutgers University, Piscataway, New Jersey 08854, United States
[‡]Department of Physics and Astronomy and CQB, Rutgers University, Piscataway, New Jersey 08854, United States
[§]Department of Mathematics and CQB, Rutgers University, Piscataway, New Jersey 08854, United States
Equal corresponding authors: E-mail: anirvans.physics@gmail.com and eduardo.sontag@gmail.com

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 Introduction

The E. coli cell-free TX-TL system (Noireaux et al. 2003; Noireaux and Libchaber 2004; Noireaux et al. 2011; Shin and Noireaux 2012; Sun et al. 2013) is a promising tool for characterizing genetic networks in a minimally complex environment, and has been proposed as a technique for large scale industrial production of bioproducts ranging from synthetic fuels to drugs [1]. This technique allows biological engineers to explore genetic circuits in an analogous way to that in which electrical engineers use breadboards to understand electronic circuits. Furthermore, the greater control provided over experimental conditions allows us to obtain a quantitative understanding of gene expression dynamics.

The experimental conditions in the TX-TL expression system must be precisely calibrated for optimal expression. Not only do various ion concentrations need to be accurately prescribed, but even resources like NTPs have to be optimized carefully for significant expression to occur. Our understanding of the resource requirements of gene expression is challenged by the observation that gene expression decreases with NTP concentration, once NTP is increased beyond a certain threshold (Siegal-Gaskins et al. 2014). This phenomenon is unexpected, since higher amounts of NTPs imply a greater source of energy available for transcription and translation. Siegal-Gaskin et al. (Siegal-Gaskins et al. 2014) speculate that increased NTP levels enhance transcriptional activity, and thus exhaust most of the present NTPs. As a result, very little resources remain for translation to proceed, explaining the reduction of gene expression. In our investigations we contrast this hypothesis with a simpler explanation involving a direct suppression of gene expression, specifically that of the translation rate as a function of the overabundance of certain reagents.

In our system, which has no significant protein degradation mechanism, conversion of ATP to ADP, and the subsequent rise of ADP/ATP ratio (Caschera and Noireaux 2014), slow down protein production rates. A few hours into the experiment, this slowdown causes protein levels to saturate. Analyzing the gene expression profile thus requires us to be cognizant of the temporal dependence of transcription and, particularly, translation rates. We have developed methods of extracting such time-dependent rates.

One could reasonably believe that at times prior to saturation, resource limitation is peripheral to the underlying dynamics of gene expression. Therefore monitoring the protein and RNA levels at early times as a function of NTP and other reagents allows us to differentiate protein level suppression at high NTP concentrations. In general, we will use this experimental strategy to explore the landscape of resource dependence in the TX-TL system.

2 Materials and methods

2.1 Cell-free reaction conditions

Crude extract was prepared from BL21-Rosetta 2 strain, following the protocol of Sun et al. (Sun et al. 2013). Basic Cell-Free reaction consists of a crude cell extract from E. coli (which contains endogenous transcription-translation machinery, mRNA and protein degradation enzymes plus all the soluble proteins), buffer and DNA. The amino acid mix was prepared following the protocol from (Caschera and Noireaux 2015). All the experiments shown here are performed using the same batch of extract preparation to avoid variation between different batches. Reactions were carried out in 10 L volume using 1 to 2 nM plasmid, 10 to 20 uM malachite green to simultaneously monitor the RNA production. The reaction temperature was set to 29° C. Experiments were run for 14 to 20 hours with readings taken every 20 minutes, so that GFP and RNA expression dynamics could be monitored over a long period of time.

2.2 Monitoring RNA production

mRNA production was monitored using a plasmid provided by Dr. Murrays lab which consists of a 35 base (MGapt) aptamer sequence incorporated in the 3 UTR of the fluorescent protein reporter

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

gene, 15 bases downstream of the stop codon and capable of binding the dye malachite green and produces fluorescence signal which can be measured at 650 nm wavelength. Fluorescence measurements were made in a total of 6 wells for each condition in a Biotek plate reader at 20 min intervals using excitation/emission wavelengths set at 610/650 nm (MGapt) and 485/525 nm (deGFP). Error bars indicate standard error over 6 replicates. The reported protein production at each time point represents a 20 min time interval.

2.3 Plasmids and Bacterial Strains

Plasmids were a kind gift from Dr. Murrays lab. BL21-Rosetta 2 strain for making crude extract was obtained from Novagen. Plasmid DNAs used for measuring protein and RNA production were prepared using Qiagen Plasmid Midi prep kits.

2.4 mRNA Preparation

First PCR was performed on pIVEX2.3 PT7-deGFP-MGapt plasmid using T7 promoter and T7 terminator primers obtained from IDT. The linear PCR was used as starting material for setting up transcription reactions using T7 RNA polymerase (Cellscript).

3 Results and discussion

We begin by examining the time-dependent gene expression profiles for increasing NTP concentrations. We use a plasmid described in Siegal-Gaskin et al., kindly provided to us by the Murray lab, with a built-in RNA aptamer (35-base MGapt sequence, which contains a binding pocket for malachite green (MG) dye and a fluorescent protein for accomplishing the simultaneous measurement of RNA and protein (Siegal-Gaskins et al. 2014). The result of a typical experiment is shown in Figure 1.

Figure 1 shows the time evolution of protein production as measured by GFP fluorescence, for various NTP concentrations. We illustrate only the initial 8 hours of the experiment, prior to saturation. Notice that protein production at early times is suppressed by increasing amounts of NTP. This observation conflicts with the hypothesis that resource limitation is driven by high transcriptional activity causing lower protein production. However, to further understand the roles of transcription and of translation, we proceeded to perform the experiments with the process of transcription eliminated. Below we describe these experiments.

We purified the mRNA from the plasmid mentioned above. Once more, we measure the protein expression from purified mRNA added to the reaction mixture, in the absence of DNA. Therefore, in this experiment, the only processes affecting gene expression are mRNA decay and the time-dependent translation rate. The Mg binding 35 bp aptamer allows us to measure mRNA levels while GFP fluorescence gives us access to protein abundance. The resulting mRNA and protein profiles are shown in Figure 2 and Figure 3, respectively. RNA profiles in Figure 2 show approximate exponential decay, as expected.

Using both the RNA and protein profiles, we extract time-dependent translation rates, following methods similar to those appearing in Siegal-Gaskin et al. from 2014. We assume that GFP (g) and MG (m) are related to protein (p) and RNA (r) expression as follows:

$$g(p) = \alpha p + g_0, \tag{1}$$

$$m(r) = \beta r + m_0. \tag{2}$$

Here g_0, m_0 are parameters accounting for autofluorescence, and in general vary between experiments, while α, β are proportionality constants which convert between microscopic measurements and the corresponding protein or mRNA concentrations. Importantly, we assume that and are fixed constants, independent of the experiment under consideration. Note that since RNA is unstable and undergoing

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

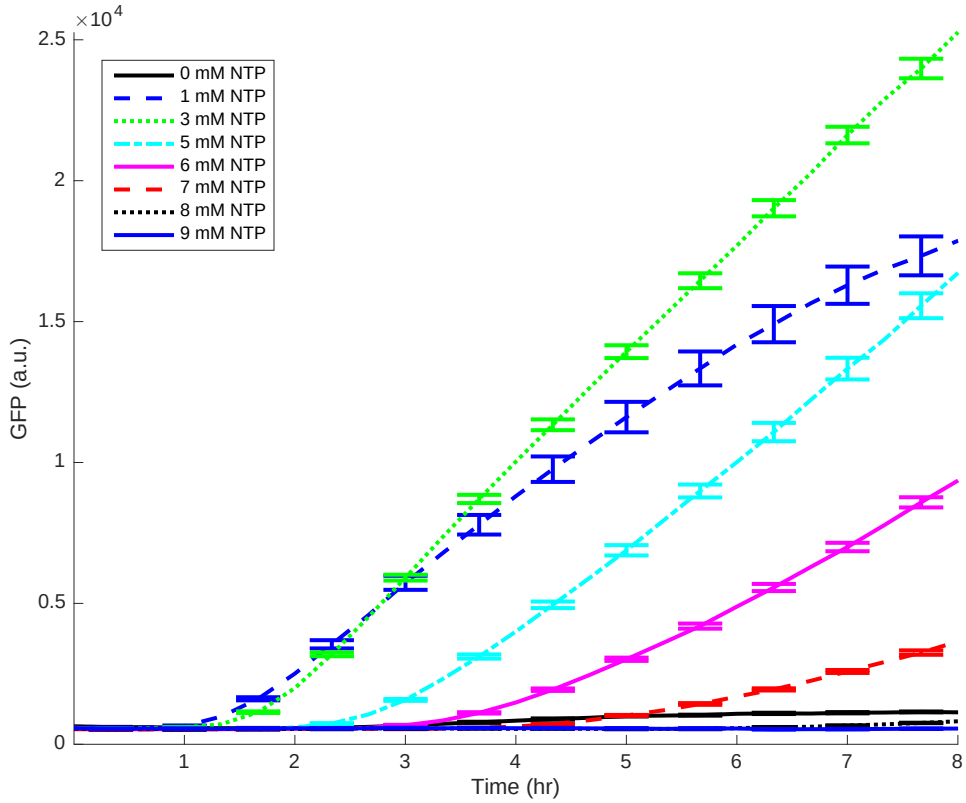

Figure 1: Translation kinetics demonstrating the GFP protein production for various NTP concentrations with DNA as the starting material. Data is measured every 20 minutes, while error bars, denoting one standard deviation, are plotted for every other data point. Concentrations of Mg and K are 17.58 mM and 91.066 mM, respectively.

approximate exponential decay (see Figure 2), we can estimate m_0 as the final temporal measurement of m for each experiment.

The fundamental quantity of interest is the (generally) time-varying translation rate $k_p(t)$. Since only translation is occurring, we assume that the rate of protein production is directly proportional to the amount of RNA $r(t)$ at time t:

$$\frac{dp}{dt}(t) \;=\; k_p(t)r(t), \tag{3}$$

To estimate $k_p(t)$, we must first relate it to the observed quantities g and m. Defining

$$M(t) \;:=\; \int_0^t (m(s) - m_0)\, ds, \tag{4}$$

$$G(M) \;:=\; g(P(M)), \tag{5}$$

one can show (for details, see the Appendix) that G(M) satisfies the following ordinary differential equation (ODE):

$$\frac{dG}{dM}(M(t)) \;=\; k_p(t). \tag{6}$$

Since and are constant across experiments, equation (5) implies that $\frac{dG}{dM}$ can be used as a proxy for the

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIG HERE

Figure 2: RNA decay profile for various NTP concentrations as a function of time, with mRNA as the starting material. Data is measured every 20 minutes, while error bars, denoting one standard deviation, are plotted for every other data point. Concentrations of Mg and K are 17.58 mM and 91.066 mM, respectively.

FIG HERE

Figure 3: Translation kinetics showing the GFP protein production for various NTP concentrations with mRNA as the starting material. Data is measured every 20 minutes, while error bars, denoting one standard deviation, are plotted for every other data point. Concentrations of Mg and K are 17.58 mM and 91.066 mM, respectively.

translation rate k_p.Thus, to analyze the qualitative effects of $Mg2^+$ and NTP on translation, we compute slopes in the G-M plane. See Figure 4 for a sample calculation, which we describe in the following.

To calculate the translation rate (technically k_p) for each experiment, we observe the following general phenomenon. For all replicates, we observe a lag time t_0 (see for example Figure 1 or 3) where protein levels appear to be approximately constant. Assuming this is generated by the autofluorescence of GFP and thus is not due to protein production, translation has not yet begun to occur. This is verified in the $G - M$ plane as in Figure 3, where a slope of approximately zero is observed initially. Afterwards, there exists a regime where translation occurs at a constant rate, which can be taken as proportional to $k_p(t)$ via equation (5). Towards the end of each experiment, as RNA is exhausted, protein production saturates, and again the slope of $G(M)$ approaches zero. The translation rate is thus defined as the slope of $G(M)$ in the region of approximate linear growth. This behavior in the early and late regimes, taken together, implies that the slope should be largest during the period of linear growth, although some exceptions exist due to noise and the difficulty in computing derivatives with discrete data. To that end, we exclude early and late data from translation rate extraction, where these effects are most noticeable. The basis for our algorithm for computing k_p is then the following: transform to the (M, G) plane, compute moving windows of fixed length (5 data points were used, but results are robust to this value), use a least-squares regression to fit a line to the data, and measure the translation rate k_p, as the maximum value over all such moving windows.

To verify that translation occurs, at least on an intermediate time window, at an approximate constant rate, we plot the GFP-MG (i.e. protein-RNA) translation data in the (G, M) plane for each experiment (each NTP and $Mg2^+$ is performed with 6 replicates). For example, see the plot in Figure 4. Blue crosses represent experimental data points. In each case, we observe rapid linear growth between regions of slower growth; the initial slow growth corresponds to the experimental delay, which implies approximately no change in protein production: $\frac{dG}{dM} = 0$, while the latter to the diminished RNA concentration in the well (i.e. saturation). This approximately linear regime is where we measure the translation rate k_p via a least-squares regression (slope of black line). Note that the clustering of data points for large M values is due to the fact that the transformation $t \to M(t)$ is not an isometry. As time t increases, the decay of RNA (Figure 2) implies that the total amount of RNA (measured via M) stabilizes.

FIG HERE

Figure 4: The change in protein vs. the integral of the RNA profile at 20.58 mM $Mg2^+$ and no additional extra NTP. For most replicates and Mg/NTP concentrations, we observe a clear region of nearly constant translation rate. The extracted rate (i.e. slope) for this experiment is shown via the black solid line.

The computed translation rate is plotted against both NTP and $Mg2^+$ concentrations in Figure 5, confirming our hypothesis that additional NTP is directly inhibiting translation. Here error bars are provided at each experimental measurement, and indicate one standard deviation of the sample mean.

Analyzing Figure 5, we observe a complex interdependence between $Mg2^+$ and NTP concentrations on the translation rate dynamics. Specifically, for each value of NTP, two distinct maxima exist as a function

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIG HERE

Figure 5: Surface plot showing the translation rate as a function of various NTP and $Mg2^+$ concentrations.

of $Mg2^+$. As the NTP concentration increases, one of these maxima shifts to increasing values of $Mg2^+$. Recalling that $Mg2^+$-NTP is the biologically relevant complex, we have developed a phenomenological model to explain both phenomenona : the shifting of a maximum, as well as the presence of multiple (here two) local maxima. Each phenomenon is described separately. We begin with a discussion of the shift of a (single) maximum. For the discussion of multiple maxima, see below.

In this model, a greater amount of NTP requires a greater amount of $Mg2^+$ to be functional. Otherwise, the free NTP, unbound to $Mg2^+$, poisons the translation system. Initially, this may appear unintuitive in our experimental system, since the total concentration of NTP (a few mMs) is 2-3 times smaller than the total $Mg2^+$ concentrations, with dissociation constant on the order of 0.01mM (Rose 1968). However, we recall that translation is affected not by the total ion concentration but instead by the free ions in the solution. Nierhaus et al. have shown (Nierhaus 2014) the binding of ribosomes to a significant fraction of $Mg2^+$ and Potassium, thereby leaving very little cytosolic ions available. The concentration of Mg available for binding with NTP to form NTP-Mg complex, which is central to gene expression (Wolf and Cittadini 2003), is therefore treated as an effective total concentration in our mathematical model.

More precisely, we assume that the translation rate takes the following general form:

$$k_p \;=\; \alpha \cdot \frac{[NTP - Mg]}{[NTP - Mg] + \beta[NTP] + \gamma} \cdot \frac{1}{1 + [Mg]_T} \tag{7}$$

Here [NTP-Mg] denotes the concentration of the NTP-Mg complex, and is the driver of translation. The above expression represents the competition at the binding site between the unbound NTP and NTP-Mg molecules, with β and γ denoting (relative) dissociation constants, and α the maximum possible translation rate. We also have included a total Mg dependence of the form 1/(1+[Mg]T), since it is known that $Mg2^+$ in abundance will inhibit translation.

To calculate the concentrations used in the above formula, we assume all reactions are in equilibrium, and thus concentrations can be calculated via a steady-state analysis. We have the following association-dissociation reaction between Mg and NTP:

$$[Mg] + [NTP]k1k - 1[NTP - Mg],$$

from which we can calculate the concentration of [NTP-Mg] as a function of [Mg]T, [NTP]T. and the dissociation constant kd (= k-1/k1). In general, we can show that for each total concentration of NTP, the above will possess a unique relative maximum, which increases as the total amount of NTP is increased. A sample plot is included in Figure 6 for demonstrative purposes.

FIG HERE

Figure 6: Heat map showing the translation rate as a function of both NTP and $Mg2^+$for the one maximum model **FIX THIS REFERENCE: [7]**.

The previous model does not capture the two-peak structure of the datasets. We believe that there are two different kinds of influence of Mg on translation. One of these has to do with NTP titration, as described before. The other has to do with stabilizing the structures of RNA and other molecules involved in the process (Lightfoot 1988; Nierhaus 2014), and that effect has an optimal concentration requirement. The superposition of two such effects would then lead to two local maxima. Disentangling these two effects from the experimental data is indeed an interesting question.

Basic cellular processes like transcription and translation require many resources. As we try to build synthetic gene regulation networks, we need to understand competition for resources giving rise to limitations and tradeoffs. Both mathematical modeling and experimental studies are directed towards

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

elucidation of this issue (Borkowski et al. 2016; Gyorgy and Del Vecchio 2014; Siegal-Gaskins et al. 2014). In the system studied, we provide evidence that suppression of gene expression due to added NTP is not primarily due to competition between transcription and translation as has been proposed in (Siegal-Gaskins et al. 2014), but we suggest that the limitation of available free $Mg2^+$ plays a role in this suppression. This suggestion is based on the observation that the suppression due to additional NTP is partly relieved by adding more $Mg2^+$. In general, because of the nontrivial interaction between different resources, more studies need be done scanning multiple parameters simultaneously, while monitoring a particular phenotype.

Lastly, the fine-tuning needed in the TX-TL raises questions regarding the optimization of gene expression in vivo. Our knowledge of sensing and controlling levels of crucial reagents in the cell is currently incomplete at best, although quite a bit is known about homeostasis of ATP levels (Hara and Kondo 2015). These synthetic biology studies therefore provide an important impetus for furthering our understanding.

Author Contributions

Conceived of the project (E.D.S.), designed the experiments and analysis (E.D.S., A.M.S, and V.H.N.), performed the experiments and did preliminary data exploration (V.H.N.), analyzed data, especially, translation rate extraction (J.M.G.), performed mathematical modeling (J.M.G) and wrote the manuscript (V.H.N., J.M.G., A.M.S and E.D.S.).

Acknowledgements

The authors thank Vincent Noireaux and members of the Murray lab for useful discussions and materials. This research is funded in part by grant ONR N00014-13-1-0074.

Data availability

All raw data are available with the online version of this article.

Supplementary data

Supplementary data are available at Synthetic Biology online.

Conflict of interest:

None declared.

4 References

References

[1] Biotechnology: Cell-free biotech will make better products. *The Economist*, May 6, 2017 2017.

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Borkowski, O., et al. (2016), 'Overloaded and stressed: whole-cell considerations for bacterial synthetic biology', Curr Opin Microbiol, 33, 123-30.

Caschera, F. and Noireaux, V. (2014), 'Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system',Biochimie, 99, 162-8.

— (2015), 'Preparation of amino acid mixtures for cell-free expression systems', Biotechniques, 58 (1), 40-3.

Gyorgy, Andras and Del Vecchio, Domitilla (2014), 'Limitations and trade-offs in gene expression due to competition for shared cellular resources', Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on (IEEE), 5431-36.

Hara, K. Y. and Kondo, A. (2015), 'ATP regulation in bioproduction', Microb Cell Fact, 14, 198.

Lightfoot, D. A. (1988), 'Magnesium-dependence of in vitro translation programmed by gene-specific mRNAs', Nucleic Acids Res, 16 (9), 4164.

Nierhaus, K. H. (2014), 'Mg2$^+$, K+, and the ribosome', J Bacteriol, 196 (22), 3817-9.

Noireaux, V. and Libchaber, A. (2004), 'A vesicle bioreactor as a step toward an artificial cell assembly', Proc Natl Acad Sci U S A, 101 (51), 17669-74.

Noireaux, V., Bar-Ziv, R., and Libchaber, A. (2003), 'Principles of cell-free genetic circuit assembly', Proc Natl Acad Sci U S A, 100 (22), 12672-7.

Noireaux, V., Maeda, Y. T., and Libchaber, A. (2011), 'Development of an artificial cell, from self-organization to computation and self-reproduction', Proc Natl Acad Sci U S A, 108 (9), 3473-80.

Rose, I. A. (1968), 'The state of magnesium in cells as estimated from the adenylate kinase equilibrium', Proc Natl Acad Sci U S A, 61 (3), 1079-86.

Shin, J. and Noireaux, V. (2012), 'An E. coli cell-free expression toolbox: application to synthetic gene circuits and artificial cells', ACS Synth Biol, 1 (1), 29-41.

Siegal-Gaskins, D., et al. (2014), 'Gene circuit performance characterization and resource usage in a cell-free "breadboard"', ACS Synth Biol, 3 (6), 416-25.

Sun, Z. Z., et al. (2013), 'Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology', J Vis Exp, (79), e50762.

Wolf, F. I. and Cittadini, A. (2003), 'Chemistry and biochemistry of magnesium', Mol Aspects Med, 24 (1-3), 3-9.

Appendix

In this section, we formally derive equation (5). Assume that the RNA concentration r(t) ¿ 0 for all t in each experiment, and that translation can be described by (3):

$$\frac{dp}{dt}(t) \;=\; k_p(t)r(t).$$

Equations (2) and (4) imply that

$$M(t) \;=\; \beta \int_0^t r(s)\,ds,$$

so that $M'(t) = r(t) > 0$,i.e. M is an increasing function of t. Thus, $M(t)$ is invertible, and we can solve

$$t \;=\; t(M)$$

on some interval $M \in [0, M_{\max})$. Note that $M_{\max} < \infty$ as $r(t)$ is exponentially decreasing, although this is not strictly necessary. Furthermore, we can compute the derivative of this function via the Inverse Function Theorem::

$$dt/dM \;=\; 1/M'(t(M)) \;=\; 1/\beta r(t(M)). \tag{8}$$

bioRxiv preprint doi: https://doi.org/10.1101/142869; this version posted May 26, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Defining the GFP marker as a function M,

$$G(M) \;:=\; g(p(t(M))),$$

is thus well-defined, and we can compute its derivative using the chain-rule:

$$\frac{dG}{dM} \;=\; \frac{dg}{dt}(p(t(M)))\frac{dp}{dt}(t(M))\frac{dt}{dM}(M).$$

Equation (1) implies that $\frac{dg}{dt} = \alpha$, equation (3) that $\frac{dp}{dt} = k_p(t(M))r(t(M))$, and equation (8) provides the last term on the right-hand side. Canceling $r(t(M))$ yields the expression

$$\frac{dG}{dM}(M) \;=\; k_p(t(M)),$$

which is precisely (6), as desired.