Modelling essential interactions between synthetic genes and their chassis cell

R. J. R. Algar, T. Ellis, G.-B. Stan

***Abstract*— In this paper we develop a model of gene expression for the purpose of studying the resource use of synthetic gene circuits when expressed in chassis cells. This model focuses on the translational aspect of gene expression which accounts for the majority of resource usage. The model allows for simulations of circuit-chassis interactions that can be used to inform improved designs of synthetic gene circuits.**

***Index Terms*— Synthetic Biology, Circuit-Chassis Interactions, Ribosomal Availability, Translational Model**

I. Background

Ideally, a cell should function as a context-free "chassis" for the synthetic biology constructs that it hosts. In practice, this is not the case as endogenous circuitry interacts with heterologous systems leading to non-linear outcomes [1], [2] even when components "orthogonal" to the host are employed. A fundamental system-level interaction is the depletion of cellular resources, such as ATP, polymerases and ribosomes, which are all required for operation of both synthetic constructs and their host cells. The resulting phenomenon, called "burden" or "load", typically sees a reduction in cell growth rate and fitness, as well as deleterious consequences for hosted synthetic circuits [3]. Numerous models have been proposed to capture the interactions between synthetic circuits and their host cells through shared resource pools [4], [5], [6], [7], [8], [9], [10], [11], and reported experimental results indicate that the major bottleneck in resource use is the availability of free ribosomes [12], [13].

To capture the major fundamental effects of circuit-chassis interactions, we therefore chose to build a translation elongation model that can be used to predict ribosome availability for synthetic biology circuits of different designs. The model enables simulations of the burden imposed on constitutive gene expression when a synthetic biology circuit in a growing cell is being "switched on", and also indicates synthetic construct design changes that can minimise circuit load.

II. Full elongation model

To create a model that captures essential features, such as the ability for protein coding sequences to have different codon profiles and the impact of codon profiles on the translation dynamics including ribosomal traffic jams [14], [15], [16], we have to look closely at the elongation process that occurs during mRNA translation. This is a complex process that consists of multiple steps every time the polypeptide chain is elongated and the ribosome moves along the mRNA transcript. Since obtaining the values for the parameters associated with these individual processes is typically very hard we decided to collapse them down such that each time the ribosome moves one codon along the transcript this is captured as a single-step process.

Centre for Synthetic Biology and Innovation and Department of Bioengineering, Imperial College London, United Kingdom. Email: {g.stan, t.ellis}@imperial.ac.uk.

In Sections II-A.1-II-A.3, we derive our model using a random-walk approach in a similar way to early work presented in [17]. We then approximate this probabilistic model by a deterministic nonlinear ODE model which can be proven to have a unique equilibrium point solution. In Section III, we confirm that our proposed model makes sense by instantiating it with parameters found in the literature and showing that the model outputs qualitatively reflect what would be expected *in vivo*. We also show how this model can be used to simulate the effects of changing various control points such as promoter strength, transcript copy numbers, RBS strength or codon usage (see [18]). Growth rate is not included in our model as the interaction between resource availability and growth rate is still an open and debated question.

A. Model derivation

In our model the movement of individual ribosomes is treated as a random walk on mRNA transcripts, which occurs based on the following assumptions:

1) Assumptions:

Assumption 1: There is a fixed total number of ribosomes R.
Assumption 2: There is a single species of transcripts of which there is a constant number M.
Assumption 3: Each transcript is identical and is of length L codons.
Assumption 4: 'Free' ribosomes can reversibly bind to the RBS of each transcript.
Assumption 5: Once elongation is initiated and a ribosome has moved to the first codon of the transcript, it must continue unidirectionally along the transcript until it reaches the stop codon.
Assumption 6: When a ribosome reaches the stop codon it will release it, become a 'free ribosome', and a protein will be produced.
Assumption 7: We approximate the size of ribosomes to be such that they only occupy a single codon (or RBS) along a transcript, and neighbouring codons (and RBSs) can be occupied by separate ribosomes.
Assumption 8: No two ribosomes can occupy the same codon or RBS at the same time.
Assumption 9: Ribosomes move along transcripts one codon at a time and cannot move to the next codon if it is occupied by another ribosome.
Assumption 10: Ribosomes move from one codon to the next only if the next codon is not occupied. This process occurs at a fixed and constant rate defined by the specific codon the ribosome is currently bound to.
Assumption 11: There is a *large* number of total ribosomes, so $R \gg 1$.
Assumption 12: Transitions from one elongation state to the next are single steps and time is modelled discretely with intervals δt. A maximum of one state transition for each ribosome may occur during this time interval δt (i.e.

maximum one elongation step per ribosome during the time interval δt).

Assumption 13: All mRNA transcripts are identical and so the probability of a ribosome r being in elongation stage $i \in \{0, ..., L\}$ at time t is the same for any mRNA transcript:

$$\mathbb{P}(E^r_{m,i}, t) = \mathbb{P}(E^r_{s,i}, t) \qquad \forall \ m, s \in \{1, ..., M\}$$

where the events notation $E^r_{m,i}$ is defined in the *"Events"* section below.

Assumption 14: All ribosomes are identical and so have the same probability distribution:

$$\mathbb{P}(E^r_{m,i}, t) = \mathbb{P}(E^q_{m,i}, t) \qquad \forall \ r, q \in \{1, ..., R\}$$

where the events notation $E^r_{m,i}$ is defined in the *"Events"* section below.

Assumption 15: The event of a new transcript creation is defined as the moment an mRNA is finished being transcribed. Ribosomes move along mRNA closely following the RNA polymerase as it transcribes, and are already moving along the mRNA before it is fully transcribed [19]. Therefore upon the creation of a new mRNA we can approximate that it is already covered in ribosomes and make a quasi steady-state approximation for the transcription dynamics.

Assumption 16: The process is Markovian and at any point in time the position of one ribosome is independent of the positions of others at that point in time.

Remark 1: We acknowledge that these assumptions do not exactly represent the reality of the complex process of translation, however we make them in order to simplify the model in a way that we do not expect will affect the core behaviours of the translation dynamics. For example, this model approximates the ribosome as only taking the space of a single codon at any one time, whereas in reality this is not the case. This approximation allows us to simplify the model significantly. The loss of detail is more than made up for by the increased ease with which we can work with the model.

Remark 2: In order to better reflect biological reality when we simulate the model below, we can consider elongations steps to be occurring on groups of codons, rather than individual codons. For example, in the simulations below we treat elongation steps as encompassing 10 codons to more realistically reflect (a) the number of codons occupied by a ribosome on a transcript and (b) the maximum number of ribosomes that can be on a transcript at any point in time. If such a modification is implemented, then β_i represents the average elongation rate over 10 codons and the transcript length, L, then needs to be divided by 10. This interpretation of the model does not have any bearing on its derivation, though it is important to note when considering how it relates to biological reality.

Hereafter, we present a full derivation from first principles of a translation model. This is done in order to provide a framework that allows easy modification upon the removal of some of the assumptions listed above. For example, many circuits (such as the toggle switch [20] or repressilator [21]) display dynamic properties and Assumption 2 does not allow for this. It is possible to follow the same derivation until Assumption 2 is used, whereafter an alternative derivation can provide a model that allows for dynamic mRNA levels, and subsequently simulations can be performed from the modified equations.

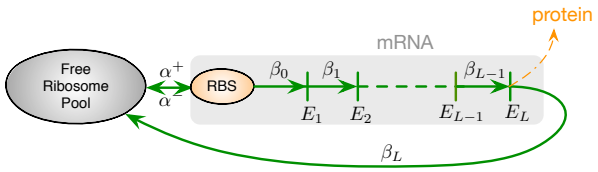

Fig. 1. Steps considered in a step-wise translation model for a single transcript. Free ribosomes bind to an unoccupied RBS at a rate α_1 and either unbind and return to the free ribosome pool at a rate α_{-1}, or elongation is initiated and the RBS-bound ribosome moves to the first position along the transcript (position E_1) at a rate β_0. Once elongation has initiated, the ribosome moves along the transcript, from position E_i to an unoccupied position E_{i+1} at a rate β_i. Once the ribosome reaches the final position E_L it returns to the free ribosome pool and a protein is produced at a rate β_L.

2) Events:

1) $E^r_{m,i}$ is the event of ribosome r being on transcript m in elongation state i (i.e. at the i^{th} codon, $i \in \{1, ..., L\}$).
2) $\neg E^r_{m,i}$ is the event of ribosome r not being on transcript m in elongation state i (i.e. at the i^{th} codon, $i \in \{1, ..., L\}$).
3) $E^r_{m,0}$ is the event of ribosome r being on the RBS of transcript m.
4) Rib^r is the event of ribosome r not being on any transcript (i.e. being in the free ribosome pool).

For any ribosome r from a pool of R ribosomes, if it is freely available (Rib^r) it can reversibly bind to the RBS of mRNA m ($E^r_{m,0}$) and, from this state, it can either unbind and join the free ribosome pool again, or translation is initiated and the ribosome then moves into the initial state of elongation ($E^r_{m,1}$). From the initial state of elongation $E^r_{m,1}$ the only path the ribosome can take is to go from the i^{th} stage of elongation ($E^r_{m,i}$) to the $i+1^{\text{th}}$ stage of elongation ($E^r_{m,i+1}$) until it reaches the final elongation stage, which, without loss of generality, can be the L^{th} stage ($E^r_{m,L}$). From this, translation finishes, a full protein is produced and the ribosome returns to the free ribosome pool. Figure 1 shows a cartoon schematic of the process we are modelling.

Notation wise, $\mathbb{P}(E^r_{m,i}, t)$ is the probability that event $E^r_{m,i}$ occurs at time t. In terms of the random walk of ribosomes on mRNA transcripts, we consider a discrete time distribution with steps of length δt and define the following 'rates':

Definition 1: We define the 'unblocked' rates:

$$\begin{aligned}
\alpha^+ &= \lim_{\delta t \to 0} \frac{\mathbb{P}\left(E^r_{m,0}, t+\delta t | Rib^r, t \cap \left(\bigcap_{q \neq r} \neg E^q_{m,0}, t+\delta t\right)\right)}{\delta t} \\
\alpha^- &= \lim_{\delta t \to 0} \frac{\mathbb{P}(Rib^r, t+\delta t | E^r_{m,0}, t)}{\delta t} \\
\beta_i &= \lim_{\delta t \to 0} \frac{\mathbb{P}\left(E^r_{m,i+1}, t+\delta t | E^r_{m,i}, t \cap \left(\bigcap_{q \neq r} \neg E^q_{m,i+1}, t+\delta t\right)\right)}{\delta t} \\
&\quad \forall i \in \{0, ..., L-1\} \\
\beta_L &= \lim_{\delta t \to 0} \frac{\mathbb{P}(Rib^r, t+\delta t | E^r_{m,L}, t)}{\delta t}
\end{aligned}$$

where α^+ is the binding rate of a ribosome to an RBS, α^- is the unbinding rate of a ribosome from an RBS and the β_i values are the rates of elongation at which a ribosome moves to the next codon (if it is not blocked).

3) Proposed translation model: The translation dynamics are expressed by the following set of equations:

$$\mathbb{P}(Rib^r, t+\delta t) = \mathbb{P}(Rib^r, t+\delta t|Rib^r, t)\mathbb{P}(Rib^r, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,0}, t)\mathbb{P}(E^r_{s,0}, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,L}, t)\mathbb{P}(E^r_{s,L}, t) \tag{1a}$$

$$\mathbb{P}(E^r_{m,0}, t+\delta t) = \mathbb{P}(E^r_{m,0}, t+\delta t|E^r_{m,0}, t)\mathbb{P}(E^r_{m,0}, t) + \mathbb{P}(E^r_{m,0}, t+\delta t|Rib^r, t)\mathbb{P}(Rib^r, t) \tag{1b}$$

$$\mathbb{P}(E^r_{m,1}, t+\delta t) = \mathbb{P}(E^r_{m,1}, t+\delta t|E^r_{m,1}, t)\mathbb{P}(E^r_{m,1}, t) + \mathbb{P}(E^r_{m,1}, t+\delta t|E^r_{m,0}, t)\mathbb{P}(E^r_{m,0}, t) \tag{1c}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,i}, t+\delta t) = \mathbb{P}(E^r_{m,i}, t+\delta t|E^r_{m,i}, t)\mathbb{P}(E^r_{m,i}, t) + \mathbb{P}(E^r_{m,i}, t+\delta t|E^r_{m,i-1}, t)\mathbb{P}(E^r_{m,i-1}, t) \quad \forall i \in \{2, ..., L-1\} \tag{1d}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,L}, t+\delta t) = \mathbb{P}(E^r_{m,L}, t+\delta t|E^r_{m,L}, t)\mathbb{P}(E^r_{m,L}, t) + \mathbb{P}(E^r_{m,L}, t+\delta t|E^r_{m,L-1}, t)\mathbb{P}(E^r_{m,L-1}, t) \tag{1e}$$

We can rewrite the probability of a ribosome staying in the same state as being equal to 1 minus the probability of it moving out of that state:

$$\mathbb{P}(Rib^r, t+\delta t) = \left(1 - \sum_{s=1}^{M} \mathbb{P}(E^r_{s,0}, t+\delta t|Rib^r, t)\right)\mathbb{P}(Rib^r, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,0}, t)\mathbb{P}(E^r_{s,0}, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,L}, t)\mathbb{P}(E^r_{s,L}, t) \tag{2a}$$

$$\mathbb{P}(E^r_{m,0}, t+\delta t) = \left(1 - \mathbb{P}(Rib^r, t+\delta t|E^r_{m,0}, t\right) - \mathbb{P}\left(E^r_{m,1}, t+\delta t|E^r_{m,0}, t\right)\right)\mathbb{P}(E^r_{m,0}, t) + \mathbb{P}(E^r_{m,0}, t+\delta t|Rib^r, t)\mathbb{P}(Rib^r, t) \tag{2b}$$

$$\mathbb{P}(E^r_{m,1}, t+\delta t) = \left(1 - \mathbb{P}(E^r_{m,2}, t+\delta t|E^r_{m,1}, t)\right)\mathbb{P}(E^r_{m,1}, t) + \mathbb{P}(E^r_{m,1}, t+\delta t|E^r_{m,0}, t)\mathbb{P}(E^r_{m,0}, t) \tag{2c}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,i}, t+\delta t) = \left(1 - \mathbb{P}(E^r_{m,i+1}, t+\delta t|E^r_{m,i}, t)\right)\mathbb{P}(E^r_{m,i}, t) + \mathbb{P}(E^r_{m,i}, t+\delta t|E^r_{m,i-1}, t)\mathbb{P}(E^r_{m,i-1}, t) \quad \forall i \in \{2, ..., L-1\} \tag{2d}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,L}, t+\delta t) = \left(1 - \mathbb{P}(Rib^r, t+\delta t|E^r_{m,L}, t)\right)\mathbb{P}(E^r_{m,L}, t) + \mathbb{P}(E^r_{m,L}, t+\delta t|E^r_{m,L-1}, t)\mathbb{P}(E^r_{m,L-1}, t) \tag{2e}$$

Rearranging then gives:

$$\mathbb{P}(Rib^r, t+\delta t) - \mathbb{P}(Rib^r, t) = -\sum_{s=1}^{M} \mathbb{P}(E^r_{s,0}, t+\delta t|Rib^r, t)\mathbb{P}(Rib^r, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,0}, t)\mathbb{P}(E^r_{s,0}, t) + \sum_{s=1}^{M} \mathbb{P}(Rib^r, t+\delta t|E^r_{s,L}, t)\mathbb{P}(E^r_{s,L}, t) \tag{3a}$$

$$\mathbb{P}(E^r_{m,0}, t+\delta t) - \mathbb{P}(E^r_{m,0}, t) = -\Big(\mathbb{P}(Rib^r, t+\delta t|E^r_{m,0}, t) + \mathbb{P}(E^r_{m,1}, t+\delta t|E^r_{m,0}, t)\Big)\mathbb{P}(E^r_{m,0}, t) + \mathbb{P}(E^r_{m,0}, t+\delta t|Rib^r, t)\mathbb{P}(Rib^r, t) \tag{3b}$$

$$\mathbb{P}(E^r_{m,1}, t+\delta t) - \mathbb{P}(E^r_{m,1}, t) = -\mathbb{P}(E^r_{m,2}, t+\delta t|E^r_{m,1}, t)\mathbb{P}(E^r_{m,1}, t) + \mathbb{P}(E^r_{m,1}, t+\delta t|E^r_{m,0}, t)\mathbb{P}(E^r_{m,0}, t) \tag{3c}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,i}, t+\delta t) - \mathbb{P}(E^r_{m,i}, t) = -\mathbb{P}(E^r_{m,i+1}, t+\delta t|E^r_{m,i}, t)\mathbb{P}(E^r_{m,i}, t) + \mathbb{P}(E^r_{m,i}, t+\delta t|E^r_{m,i-1}, t)\mathbb{P}(E^r_{m,i-1}, t) \quad \forall i \in \{2, ..., L-1\} \tag{3d}$$

$$\vdots$$

$$\mathbb{P}(E^r_{m,L}, t+\delta t) - \mathbb{P}(E^r_{m,L}, t) = -\mathbb{P}(Rib^r, t+\delta t|E^r_{m,L}, t)\mathbb{P}(E^r_{m,L}, t) + \mathbb{P}(E^r_{m,L}, t+\delta t|E^r_{m,L-1}, t)\mathbb{P}(E^r_{m,L-1}, t) \tag{3e}$$

For all events $E^r_{m,i}$ at a time t we have that the probability $\mathbb{P}(E^r_{m,i}, t|X)$ for any event X can be split into two subsets, one where there is a ribosome in elongation state i on mRNA m at time t and one where there is not:

$$\mathbb{P}(E^r_{m,i}, t|X)\mathbb{P}(X) = \mathbb{P}(E^r_{m,i}, t|X \cap (\bigcup_{q\neq r} E^q_{m,i}, t))\mathbb{P}(X \cap (\bigcup_{q\neq r} E^q_{m,i}, t)) + \mathbb{P}(E^r_{m,i}, t|X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))\mathbb{P}(X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))$$

Since the probability of two ribosomes being in the same state on the same mRNA is zero we must have that:

$$\mathbb{P}(E^r_{m,i}, t|X)\mathbb{P}(X) = \mathbb{P}(E^r_{m,i}, t|X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))\mathbb{P}(X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))$$

which can be rewritten as:

$$\mathbb{P}(E^r_{m,i}, t|X)\mathbb{P}(X) = \mathbb{P}(E^r_{m,i}, t|X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))\mathbb{P}(\bigcap_{q\neq r} \neg E^q_{m,i}, t|X)\mathbb{P}(X)$$

Due to mutual exclusivity, we know that the probability of no other ribosomes being there is equal to 1 minus the sum of the probabilities of each other ribosome being there:

$$\mathbb{P}(E^r_{m,i}, t|X)\mathbb{P}(X) = \mathbb{P}(E^r_{m,i}, t|X \cap (\bigcap_{q\neq r} \neg E^q_{m,i}, t))(1 - \sum_{q\neq r} \mathbb{P}(E^q_{m,i}, t|X))\mathbb{P}(X)$$

Combining this with equations (3) gives:

$$\mathbb{P}(Rib^r, t+\delta t) - \mathbb{P}(Rib^r, t) = -\sum_{s=1}^{M} \left(\mathbb{P}\Big(E^r_{s,0}, t+\delta t|\Big(Rib^r, t \cap (\bigcap_{q\neq r} \neg E^q_{s,0}, t+\delta t)\Big)\right)$$

$$\cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{s,0},t+\delta t|Rib^r,t)\Big)\mathbb{P}(Rib^r,t)\Big)$$
$$+\sum_{s=1}^{M}\mathbb{P}(Rib^r,t+\delta t|E^r_{s,0},t)\mathbb{P}(E^r_{s,0},t)$$
$$+\sum_{s=1}^{M}\mathbb{P}(Rib^r,t+\delta t|E^r_{s,L},t)\mathbb{P}(E^r_{s,L},t) \tag{4a}$$

$$\begin{aligned}\mathbb{P}(E^r_{m,0},t+\delta t)-\mathbb{P}(E^r_{m,0},t)=\\ -\mathbb{P}(Rib^r,t+\delta t|E^r_{m,0},t)\mathbb{P}(E^r_{m,0},t)\\ -\mathbb{P}\Big(E^r_{m,1},t+\delta t|\Big(E^r_{m,0},t\cap(\bigcap_{q\neq r}\neg E^q_{m,1},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,1},t+\delta t|E^r_{m,0},t)\Big)\mathbb{P}(E^r_{m,0},t)\\ +\mathbb{P}\Big(E^r_{m,0},t+\delta t|\Big(Rib^r,t\cap(\bigcap_{q\neq r}\neg E^q_{m,0},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,0},t+\delta t|Rib^r,t)\Big)\mathbb{P}(Rib^r,t)\end{aligned} \tag{4b}$$

$$\begin{aligned}\mathbb{P}(E^r_{m,1},t+\delta t)-\mathbb{P}(E^r_{m,1},t)=\\ -\mathbb{P}\Big(E^r_{m,2},t+\delta t|\Big(E^r_{m,1},t\cap(\bigcap_{q\neq r}\neg E^q_{m,2},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,2},t+\delta t|E^r_{m,1},t)\Big)\mathbb{P}(E^r_{m,1},t)\\ +\mathbb{P}\Big(E^r_{m,1},t+\delta t|\Big(E^r_{m,0},t\cap(\bigcap_{q\neq r}\neg E^q_{m,1},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,1},t+\delta t|E^r_{m,0},t)\Big)\mathbb{P}(E^r_{m,0},t)\end{aligned} \tag{4c}$$

$$\vdots$$

$$\begin{aligned}\mathbb{P}(E^r_{m,i},t+\delta t)-\mathbb{P}(E^r_{m,i},t)=\\ -\mathbb{P}\Big(E^r_{m,i+1},t+\delta t|\Big(E^r_{m,i},t\cap(\bigcap_{q\neq r}\neg E^q_{m,i+1},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,i+1},t+\delta t|E^r_{m,i},t)\Big)\mathbb{P}(E^r_{m,i},t)\\ +\mathbb{P}\Big(E^r_{m,i},t+\delta t|\Big(E^r_{m,i-1},t\cap(\bigcap_{q\neq r}\neg E^q_{m,i},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,i},t+\delta t|E^r_{m,i-1},t)\Big)\mathbb{P}(E^r_{m,i-1},t)\\ \forall i\in\{2,...,L-1\}\end{aligned} \tag{4d}$$

$$\vdots$$

$$\begin{aligned}\mathbb{P}(E^r_{m,L},t+\delta t)-\mathbb{P}(E^r_{m,L},t)=\\ -\mathbb{P}(Rib^r,t+\delta t|E^r_{m,L},t)\mathbb{P}(E^r_{m,L},t)\\ +\mathbb{P}\Big(E^r_{m,L},t+\delta t|\Big(E^r_{m,L-1},t\cap(\bigcap_{q\neq r}\neg E^q_{m,L},t+\delta t)\Big)\Big)\\ \cdot\Big(1-\sum_{q\neq r}\mathbb{P}(E^q_{m,L},t+\delta t|E^r_{m,L-1},t)\Big)\mathbb{P}(E^r_{m,L-1},t)\end{aligned} \tag{4e}$$

Dividing both sides by δt and taking $\lim_{\delta t\to 0}$ as well as using the rates defined above and Assumptions 1, 2, 3 and 11 gives:

$$\begin{aligned}\frac{d\mathbb{P}(Rib^r,t)}{dt} &= M\cdot\alpha^-\mathbb{P}(E^r_{m,0},t)\\ &\quad -M\cdot\alpha^+\mathbb{P}(Rib^r,t)(1-R\cdot\mathbb{P}(E^r_{m,0},t))\\ &\quad +M\cdot\beta_L\mathbb{P}(E^r_{m,L},t) \qquad (5a)\\ \frac{d\mathbb{P}(E^r_{m,0},t)}{dt} &= -\alpha^-\mathbb{P}(E^r_{m,0},t)\\ &\quad +\alpha^+\mathbb{P}(Rib^r,t)(1-R\cdot\mathbb{P}(E^r_{m,0},t))\\ &\quad -\beta_0\mathbb{P}(E^r_{m,0},t)(1-R\cdot\mathbb{P}(E^r_{m,1},t)) \qquad (5b)\\ \frac{d\mathbb{P}(E^r_{m,1},t)}{dt} &= \beta_0\mathbb{P}(E^r_{m,0},t)(1-R\cdot\mathbb{P}(E^r_{m,1},t))\\ &\quad -\beta_1\mathbb{P}(E^r_{m,1},t)(1-R\cdot\mathbb{P}(E^r_{m,2},t)) \qquad (5c)\\ &\vdots\\ \frac{d\mathbb{P}(E^r_{m,i},t)}{dt} &= \beta_{i-1}\mathbb{P}(E^r_{m,i-1},t)(1-R\cdot\mathbb{P}(E^r_{m,i},t))\\ &\quad -\beta_i\mathbb{P}(E^r_{m,i},t)(1-R\cdot\mathbb{P}(E^r_{m,i+1},t)) \qquad (5d)\\ &\vdots\\ \frac{d\mathbb{P}(E^r_{m,L},t)}{dt} &= \beta_{L-1}\mathbb{P}(E^r_{m,L-1},t)(1-R\cdot\mathbb{P}(E^r_{m,L},t))\\ &\quad -\beta_L\mathbb{P}(E^r_{m,L},t) \qquad (5e)\end{aligned}$$

A set of random variables $X_{m,i}$ ($i\in\{0...L\}$) is defined as follows:

$$X_{m,i}(t)=\begin{cases}1 & \text{if there is a ribosome present in elongation stage '}i\text{' on mRNA '}m\text{' at time '}t\text{'}\\ 0 & \text{if there is no ribosome present in elongation stage '}i\text{' on mRNA '}m\text{' at time '}t\text{'}\end{cases}$$

At any time 't', using Assumption 17 on independence of ribosome positions, we have:

$$\mathbb{P}(X_{m,i}(t)=1)=\sum_r\mathbb{P}(E^r_{m,i},t)$$

and

$$\mathbb{P}(X_{m,i}(t)=0)=1-\sum_r\mathbb{P}(E^r_{m,i},t)$$

By the definition of expectation, we then have:

$$\mathbb{E}(X_{m,i}(t))=1\cdot\mathbb{P}(X_{m,i}(t)=1)+0\cdot\mathbb{P}(X_{m,i}(t)=0)$$

so that,

$$\mathbb{E}(X_{m,i}(t))=\mathbb{P}(X_{m,i}(t)=1)$$

Using Assumption 1 we get:

$$\mathbb{E}(X_{m,i}(t))=R\cdot\mathbb{P}(E^r_{m,i},t) \tag{6}$$

where R is the total number of ribosomes. We further define the random variable $X_i(t)$ as the sum of random variables $X_{m,i}(t)$ across all mRNA, i.e. the total number of ribosomes in position i across all transcripts:

$$X_i(t)=\sum_m X_{m,i}(t)$$

which, taking expectations, gives

$$\mathbb{E}(X_i(t))=\sum_m\mathbb{E}(X_{m,i}(t))$$

Combining with equation (6) gives

$$\mathbb{E}(X_i(t)) = \sum_m R \cdot \mathbb{P}(E^r_{m,i}, t) \tag{7}$$

Now, using Assumption 2, we have:

$$\mathbb{E}(X_i(t)) = MR \cdot \mathbb{P}(E^r_{m,i}, t) \tag{8}$$

We define the variable $Y_i(t)$ to be the expectation of the random variable $X_i(t)$

$$Y_i(t) = \mathbb{E}(X_i(t)) \tag{9}$$

Therefore,

$$Y_i(t) = MR \cdot \mathbb{P}(E^r_{m,i}, t) \quad \forall i \in \{0, ..., L\} \tag{10}$$

We now investigate the variance of $X_i(t)$. The variance of each $X_{m,i}(t)$ is equal to:

$$Var(X_{m,i}(t)) \quad = \quad \mathbb{E}\left(X_{m,i}(t)^2\right) - \mathbb{E}\left(X_{m,i}(t)\right)^2 \tag{11}$$

However, since $X_{m,i}(t)$ can only take the values 0 or 1, we have that $X_{m,i}(t)^2 = X_{m,i}(t)$ and so, if we let $\mu = \mathbb{E}(X_{m,i}(t))$, we obtain:

$$Var(X_{m,i}(t)) \quad = \quad \mu - \mu^2 \tag{12}$$

$X_i(t)$ is a random variable that represents the sum of independent, identically distributed (IID) random variables $X_{m,i}(t)$. From the variance of independent, identically distributed random variables we get:

$$Var(X_i(t)) \quad = \quad \frac{\mu - \mu^2}{M} \tag{13}$$

This gives us an estimate of the cell-to-cell variance we would expect from this model. This indicates that the behaviour of the circuit becomes less noisy as the number of transcripts increases since the variance per cell is inversely proportional to the total number of transcripts. This suggests that a stronger promoter would cause lower cell-to-cell variation.

The random variable F, which represents the number of free ribosomes, can be calculated as the total number of ribosomes minus the expected total number of ribosomes on transcripts:

$$F(t) = R - \sum_i X_i(t) \tag{14}$$

Letting $G(t)$ be the expectation of the random variable $F(t)$

$$G(t) = \mathbb{E}(F(t)), \tag{15}$$

Combining (10) and (14) with (5), and dropping the (t) from the notation by letting $Y_i = Y_i(t)$ and $G = G(t)$ we are left with:

$$\frac{dG}{dt} = -M\alpha^+ G(1 - Y_0/M) + \alpha^- Y_0 + \beta_L Y_L \tag{16a}$$

$$\frac{dY_0}{dt} = M\alpha^+ G(1 - Y_0/M) - \alpha^- Y_0 - \beta_0 Y_0(1 - Y_1/M) \tag{16b}$$

$$\frac{dY_1}{dt} = \beta_0 Y_0(1 - Y_1/M) - \beta_1 Y_1(1 - Y_2/M) \tag{16c}$$

$$\vdots$$

$$\frac{dY_i}{dt} = \beta_{i-1} Y_{i-1}(1 - Y_i/M) - \beta_i Y_i(1 - Y_{i+1}/M) \tag{16d}$$

$$\vdots$$

$$\frac{dY_L}{dt} = \beta_{L-1} Y_{L-1}(1 - Y_L/M) - \beta_L Y_L \tag{16e}$$

The steady state equations, obtained assuming that the system is in exponential growth and that each transcript has a steady state distribution of ribosomes on it (Assumption 15), are then given by:

$$\begin{aligned} M\alpha^+ G(1 - Y_0/M) &= \alpha^- Y_0 + \beta_L Y_L & (17)\\ \beta_0 Y_0(1 - Y_1/M) &= \beta_1 Y_1(1 - Y_2/M) \\ &= \vdots \\ &= \beta_{i-1} Y_{i-1}(1 - Y_i/M) \\ &= \vdots \\ &= \beta_{L-1} Y_{L-1}(1 - Y_L/M) \\ &= \beta_L Y_L & (18) \end{aligned}$$

B. Solving the steady-state equations

Rearranging Equation (18) we can define functions f_G and $f_k, \forall k \in \{0, ..., L-1\}$ such that:

$$\begin{aligned} G &= f_G(Y_0, Y_1) & (19a)\\ Y_k &= f_k(Y_{k+1}, Y_{k+2}),\ \forall k \in \{0, ..., L-2\} & (19b)\\ Y_{L-1} &= f_{L-1}(Y_L) & (19c) \end{aligned}$$

These functions can be rewritten as:

$$\begin{aligned} G &= g_G(Y_L) & (20a)\\ Y_k &= g_k(Y_L),\ \forall k \in \{0, ..., L-2\} & (20b)\\ Y_{L-1} &= f_{L-1}(Y_L) & (20c) \end{aligned}$$

C. Proving uniqueness of the steady-state solution

It can be shown that g_G, f_{L-1} and g_k for $\forall k \in \{1, ..., L-2\}$ in (20) are all strictly monotonically increasing functions of Y_L and therefore (by the inverse function theorem) have inverse functions (see [18] for details). This means we can rewrite all the variables as strictly monotonically increasing functions of Rib:

$$Y_k \quad = \quad h_k(G) \qquad \forall k \in \{0, ..., L\} \tag{21a}$$

On the other hand, conservation of ribosomes imposes:

$$G + \sum_{k=0}^{L} h_k(G) = R \tag{22}$$

The left-hand-side of this equation is a sum of strictly monotonically increasing functions and therefore is itself a strictly monotonically increasing function of G which tends to $+\infty$ as G tends to $+\infty$. Therefore, for any total amount of ribosomes R, (22) has unique solution for G and therefore, the equations in (21) also have a unique solution for Y_k, $\forall k \in \{0, ..., L\}$.

III. Simulating how expression of a regulated gene impacts on the expression of an unregulated gene

Equations (21) and (22) cannot be solved analytically for systems that are large enough to be representative of realistic synthetic circuits. Therefore, the model must be numerically simulated. For this purpose, a Python script was built that allowed our proposed model to be easily simulated (see [18] for details about this Python script and its associated code).

A. Model simulations with realistic parameter values

In order to test our model, we performed numerical simulations with realistic values for *E. Coli* obtained from BioNumbers [22]. Table I shows the parameters used during our simulations. These parameters roughly correspond to a single genome-based gene with a medium promoter (2-4 transcripts per promoter in a cell at any time) giving 3 transcripts on average, and a 1500 bp CDS (500 amino acids long, or 50 codon groups) that has been codon optimised so ribosomes move at a rate of 10 codon groups per second.

Remark 3: As mentioned above, to more realistically take into account the footprint of each ribosome on an mRNA a simple modification to the above model has been made for the following simulations: an elongation step corresponds to a movement by the ribosome of 10 codons (30 nucleotides) at a time, rather than one codon at a time.

Parameter	Model Parameter	Value	Units (rib=ribosomes)
Codon speed (elongation rate)	β_i for all i	1	(group of codons) s^{-1}
Total amount of transcripts*	M	3	mRNA $cell^{-1}$
Total amount of available ribosomes**	R	10000	rib $cell^{-1}$
Transcript length	L	50	group of codons
Ribosome-RBS binding rate	α^+	0.0001	rib^{-1} RBS^{-1} s^{-1}
Ribosome-RBS unbinding rate	α^-	200	rib-RBS^{-1} s^{-1}

TABLE I

Model parameters used for testing model validity. *For a medium strength promoter (estimated from Bionumber ID 107667). **Estimate for ribosomes per cell at 37°C with doubling time of 40 mins is 26300 (Bionumber ID 102015), but in exponentially growing *E. coli* the max percentage of unnecessary protein expression is 30 to 40%[9] and therefore 38% of total ribosomes (10000) are available for this simulation.

Running a simulation with the values reported in Table I shows that the gene uses an average of 41.8 ribosomes at any point in time, which is 0.16% of all cellular ribosomes, assuming the total number of ribosomes is 26300 (Bionumber ID 102015). This appears to be the correct order of magnitude since there are approximately 4000 genes in the cell, of which around half are active in exponential growth. This gives 2000 active promoters with, on average, 1-5 mRNA transcripts per promoter per cell, giving a total of 2000-10000 mRNA transcripts per cell. The 3 mRNA transcripts from the simulated gene constitute 0.03-0.15% of the total cellular mRNAs, and therefore it appears valid that these transcripts use 0.16% of the cell's ribosomes as they have a medium-strength RBS.

B. Simulating the impact of a synthetic regulated gene on its host cell

It is trivial to extend the model described in Section II-A.3 to a system of two (or more) genes. In what follows, we performed simulations of the dynamic behaviour of a two-genes system for which one gene is a constitutively-expressed gene whose protein number can be closely monitored (called hereafter **"monitor"**) and the other gene represents an inducible synthetic gene circuit whose design and part-composition can be varied (called hereafter **"circuit"**).

Simulations of the "monitor-circuit" system were performed to obtain predictions about changes in the behaviour of a synthetic circuit interacting with the free ribosome pool of its host cell. During all simulations the monitor parameters are kept constant while those of the circuit are varied. In particular, we ran simulations to obtain predictions of the monitor-circuit dynamic interaction resulting from changing some control parts of the circuit such as promoter strength and plasmid copy number, or RBS strength and codon usage, and see how these affect the circuit and monitor behaviours.

1) Promoter strength and copy number: The model being considered in this paper only captures ribosomal availability and therefore when considering the number of circuit transcripts, it is independent of the mechanisms that cause changes in the amount of mRNA transcripts. In this modelling approach the plasmid copy number and promoter strength are compounded into a single variable: the number of transcripts. Figure 2 shows the amount of circuit output and monitor output for a range of transcript numbers. At low levels of transcripts (< 400 per cell) the relationship between transcript number and circuit output is approximately proportional. Similarly, the relationship between the number of transcripts and monitor output is approximately linear in this region. This indicates that for a given number of ribosomes and for transcript numbers in this range, all transcripts use a similar number of ribosomes to produce proteins at a similar rate.

As the number of transcripts increases, the system becomes saturated with respect to transcripts and large increases in the number of transcripts cause relatively small increases (resp. decreases) in circuit output (resp. monitor output).

The vertical yellow lines in Figure 2 show the time points corresponding to the simulated data given in Figure 3.

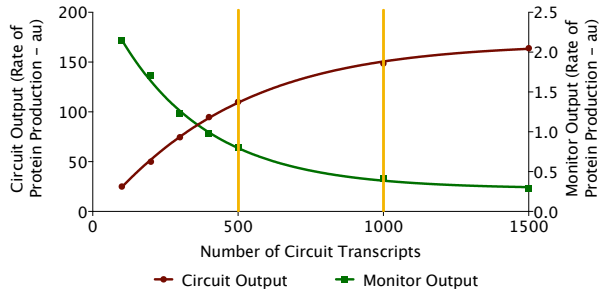

Fig. 2. Modelled impact of transcript number on circuit and monitor outputs. This figure shows both monitor (green) and circuit (red) outputs for a range of circuit transcript numbers. Plain curves represent best fit of Michaelis-Menten curves.

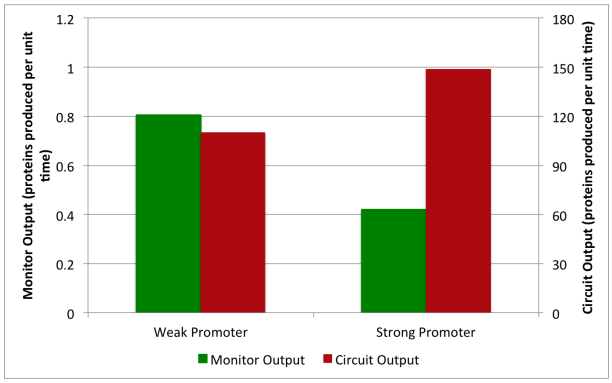

Fig. 3. Modelled impact of transcript number on circuit (red) and monitor output (green). This figure shows that a higher number of transcripts causes a higher circuit output and a lower monitor output.

2) RBS strength and codon usage: We modelled the system with a range of different RBS strengths as well as two different codons usages. The fast codon version has uniform elongation rates of 1 along a transcript composed of 100 codons and the slow codon version has uniform elongation rates of 1 along a 100 codon transcript with the exception of elongation rates of 0.5 for codons 85 to 95.

Both codon usage and RBS have a large impact on the behaviour of the circuit. Figure 4 shows how both codon usage and RBS strength affect the monitor output and the circuit output. For both codon usages, as the RBS strength increases at low levels (< 0.4) the relationship between circuit output and RBS strength is approximately linear. As the RBS strength continues to increase, the circuit output reaches a saturation level. Slower codons heavily impact the maximum output of the circuit. This is due to slower codons imposing a lower maximum flux of ribosomes through the system. Also, for slower codons this saturation is reached at lower RBS strength. This intuitively makes sense since slower codons will cause a lower maximum flux and a lower rate of recruitment of ribosomes onto the transcript will cause this maximum to be reached.

In terms of monitor output, for RBS strengths < 1, the relationship between RBS strength and monitor output is approximately linear. For higher RBS strengths, the monitor output tends to a lower asymptote. The slower codon circuit causes a decrease in monitor output (see Fig. 4).

The vertical yellow lines in Fig. 4 represent the RBS strengths (RBS strengths of 0.3 and 2 respectively) at which the data represented in Fig. 5 are considered. The dashed blue line represents the RBS strength (value = 1) at which the data represented in Fig. 6 are considered.

Remark 4: Note that our model is unable to capture the known phenomenon of reduced circuit output at the highest RBS strengths. This is because we are not including cellular response and adaptation in this model. Such considerations are beyond the scope of this paper and we plan to incorporate these in future work.

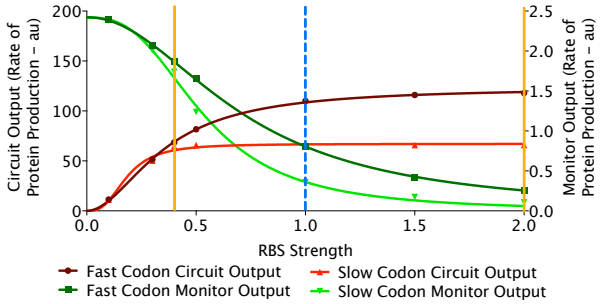

Fig. 4. Modelled impact of RBS strength and codon usage on circuit and monitor outputs. The figure shows both monitor output (green colours) and circuit output (red colours) for a range of RBS strengths for two different codon usages. Lines represent best fit of hill curves. The two vertical yellow lines represent the RBS strengths considered in Figure 5 while the dashed blue line represents the RBS strength considered in Figure 6.

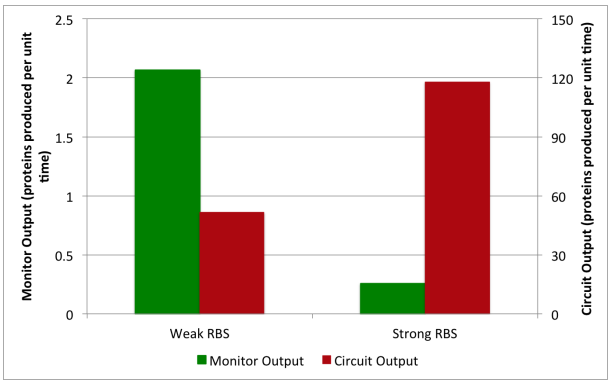

Fig. 5. Modelled impact of RBS strength on circuit and monitor output in the fast codon case. These simulation data show that a stronger RBS causes higher circuit output and lower monitor output.

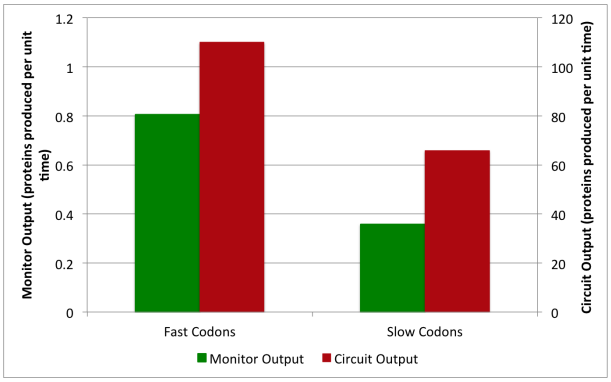

Fig. 6. Modelled impact of codon usage on circuit and monitor output shows that slower codons in the circuit cause lower circuit output as well as lower monitor output.

IV. Conclusion

In this paper we have shown the development of a model of translation elongation. This model has been designed so that it is able to incorporate the effect of codon usage on gene expression as well as ribosomal usage. This model assumes that the competition for transcriptional resources is (much) less important than the competition for translational resources.

To develop our model, we used a random walk approach to capture the dynamic behaviour within a cell of free ribosomes reversibly binding to an mRNA transcript before moving unidirectionally along the transcript. Using expectations, we then approximated this probabilistic model by a deterministic nonlinear ODE model which can be proven to have a unique equilibrium point solution. This nonlinear model was numerically simulated using a Python script that is able to simulate a cell with an arbitrary number of mRNA species where the length, codon speed, RBS strength and number of transcripts can all be defined. Using this Python script we ran simulations of our proposed nonlinear model with biologically realistic numbers and obtained outputs that were within realistic bounds with respect to values reported in the literature.

We further used this model to simulate the impact of changing the number of transcripts (to reflect a change in copy number or promoter strength), as well as RBS strength and codon usage. These simulation results showed that there are diminishing returns for protein production levels as transcript numbers or RBS strengths are increased, or when slow codons are introduced. Overall, this approach allows us to better understand the dynamic relationship that exists, through shared ribosomes, between constitutive and regulated genes that are co-expressed within a cell.

Finally, we have assessed experimentally our proposed translation model by constructing a genome-based monitor in *E. coli* and introducing plasmid-borne circuits with different copy numbers, promoter strengths, RBS strengths and codon usages. The collected experimental data confirm the results obtained through this theoretical study and are presented in a separate paper that is currently under review.

References

[1] S. Cardinale and A. Arkin, "Contextualizing context for synthetic biology–identifying causes of failure of synthetic biological systems." *Biotechnology journal*, vol. 7, no. 7, pp. 856–866, 2012. [Online]. Available: http://dx.doi.org/10.1002/biot.201200085

[2] D. Del Vecchio, A. Ninfa, and E. Sontag, "Modular cell biology: retroactivity and insulation." *Molecular systems biology*, vol. 4, 2008. [Online]. Available: http://dx.doi.org/10.1038/msb4100204

[3] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala, "Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria," *Biotechnology And Bioengineering*, vol. 35, no. 7, pp. 668–681, 1990.

[4] S. Peretti and J. Bailey, "Mechanistically detailed model of cellular metabolism for glucose-limited growth of escherichia coli B/r-A." *Biotechnology and bioengineering*, vol. 28, no. 11, pp. 1672–1689, 1986. [Online]. Available: http://dx.doi.org/10.1002/bit.260281111

[5] ——, "Simulations of host-plasmid interactions in escherichia coli: Copy number, promoter strength, and ribosome binding site strength effects on metabolic activity and plasmid gene expression." *Biotechnology and bioengineering*, vol. 29, no. 3, pp. 316–328, 1987. [Online]. Available: http://dx.doi.org/10.1002/bit.260290305

[6] M. Marchisio and J. Stelling, "Computational design of synthetic gene circuits with composable parts." *Bioinformatics (Oxford, England)*, vol. 24, no. 17, pp. 1903–1910, 2008. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/btn330

[7] S. Klumpp, Z. Zhang, and T. Hwa, "Growth rate-dependent global effects on gene expression in bacteria." *Cell*, vol. 139, no. 7, pp. 1366–1375, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.cell.2009.12.001

[8] S. Levy and N. Barkai, "Coordination of gene expression with growth rate: a feedback or a feed-forward strategy?" *FEBS letters*, vol. 583, no. 24, pp. 3974–3978, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.febslet.2009.10.071

[9] M. Scott, C. Gunderson, E. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of cell growth and gene expression: origins and consequences." *Science (New York, N.Y.)*, vol. 330, no. 6007, pp. 1099–1102, 2010. [Online]. Available: http://dx.doi.org/10.1126/science.1192588

[10] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppin, and T. Tuller, "Genome-scale analysis of translation elongation with a ribosome flow model." *PLoS computational biology*, vol. 7, no. 9, p. e1002127, Sep. 2011.

[11] Y.-B. Zhao and J. Krishnan, "mRNA translation and protein synthesis: an analysis of different modelling methodologies and a new PBN based approach." *BMC systems biology*, vol. 8, no. 1, p. 25, Feb. 2014.

[12] J. Vind, M. Sorensen, M. Rasmussen, and S. Pedersen, "Synthesis of proteins in escherichia coli is limited by the concentration of free ribosomes. expression from reporter genes does not always reflect functional mRNA levels." *Journal of molecular biology*, vol. 231, no. 3, pp. 678–688, 1993. [Online]. Available: http://dx.doi.org/10.1006/jmbi.1993.1319

[13] D. De Vos, F. Bruggeman, H. Westerhoff, and B. Bakker, "How molecular competition influences fluxes in gene expression networks." *PloS one*, vol. 6, no. 12, 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0028494

[14] T. Chou and G. Lakatos, "Clustered bottlenecks in mRNA translation and protein synthesis." *Physical review letters*, vol. 93, no. 19, 2004.

[15] A. Basu and D. Chowdhury, "Traffic of interacting ribosomes: Effects of single-machine mechanochemistry on protein synthesis," *Physical Review E*, vol. 75, 2007. [Online]. Available: http://dx.doi.org/10.1103/PhysRevE.75.021902

[16] N. Mitarai, K. Sneppen, and S. Pedersen, "Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization." *Journal of molecular biology*, vol. 382, no. 1, pp. 236–245, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.jmb.2008.06.068

[17] C. T. MacDonald and J. H. Gibbs, "Concerning the kinetics of polypeptide synthesis on polyribosomes," *Biopolymers*, vol. 7, 1969. [Online]. Available: http://dx.doi.org/10.1002/bip.1969.360070508

[18] R. Algar, T. Ellis, and G.-B. Stan, "Modelling the burden caused by gene expression: an in silico investigation into the interactions between synthetic gene circuits and their chassis cell," *arXiv.org*, Sep. 2013.

[19] S. Proshkin, A. Rahmouni, A. Mironov, and E. Nudler, "Cooperation between translating ribosomes and RNA polymerase in transcription elongation." *Science (New York, N.Y.)*, vol. 328, no. 5977, pp. 504–508, 2010. [Online]. Available: http://dx.doi.org/10.1126/science.1184939

[20] T. Gardner, C. Cantor, and J. Collins, "Construction of a genetic toggle switch in escherichia coli." *Nature*, vol. 403, no. 6767, pp. 339–342, 2000. [Online]. Available: http://dx.doi.org/10.1038/35002131

[21] M. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators." *Nature*, vol. 403, no. 6767, pp. 335–338, 2000. [Online]. Available: http://dx.doi.org/10.1038/35002125

[22] "Bionumbers." [Online]. Available: http://bionumbers.hms.harvard.edu/