

SCIENTIFIC REPORTS



OPEN

Optimal Down Regulation of mRNA Translation

Yoram Zarai¹, Michael Margaliot^{1,2} & Tamir Tuller^{2,3}

Received: 10 May 2016
Accepted: 19 December 2016
Published: 25 January 2017

Down regulation of mRNA translation is an important problem in various bio-medical domains ranging from developing effective medicines for tumors and for viral diseases to developing attenuated virus strains that can be used for vaccination. Here, we study the problem of down regulation of mRNA translation using a mathematical model called the ribosome flow model (RFM). In the RFM, the mRNA molecule is modeled as a chain of n sites. The flow of ribosomes between consecutive sites is regulated by $n + 1$ transition rates. Given a set of feasible transition rates, that models the outcome of all possible mutations, we consider the problem of maximally down regulating protein production by altering the rates within this set of feasible rates. Under certain conditions on the feasible set, we show that an optimal solution can be determined efficiently. We also rigorously analyze two special cases of the down regulation optimization problem. Our results suggest that one must focus on the position along the mRNA molecule where the transition rate has the strongest effect on the protein production rate. However, this rate is not necessarily the slowest transition rate along the mRNA molecule. We discuss some of the biological implications of these results.

Gene expression is the process by which the genetic code inscribed in the DNA is transformed into proteins. The process consists of four main steps: *transcription* of a DNA gene into an mRNA molecule, *translation* of the mRNA molecule to a protein, degradation of mRNA molecules, and degradation of proteins. During mRNA translation, macromolecules called ribosomes move unidirectionally along the mRNA molecule, decoding it codon by codon into a corresponding chain of amino acids that is folded to become a functional protein. Translation is a fundamental biological process, and understanding and re-engineering this process is important in many scientific disciplines including medicine, evolutionary biology, and synthetic biology¹.

New methods that measure gene-specific translation activity at the whole-genome scale, like polysome profiling² and ribosome profiling³, have led to a growing interest in mathematical models for translation. Such models can be used to integrate and explain the rapidly accumulating biological data as well as to predict the outcome of various manipulations of the genetic machinery. Recent methods that allow *real-time imaging* of translation on a single mRNA transcript *in vivo* (see, e.g. refs 4–7) are expected to provide even more motivation for developing and analyzing powerful dynamical models of translation.

Down-regulation of translation is important in cell biology, medicine, and biotechnology. Indeed, diverse viruses down regulate host translation by cleaving proteins such as eIF4G⁸ in order to free more resources for translating the viral genes. As another example, in many organisms small RNA genes, such as microRNAs, hybridize to the mRNA in specific locations^{9,10} in order to down-regulate translation initiation or elongation^{11,12} and/or promote mRNA degradation. Alterations in the expression of microRNA genes contribute to the pathogenesis of most, if not all, human malignancies¹³, and many times cancer cells are targeted via generating tumor specific RNA interference (RNAi) genes that down-regulate the oncogenes^{14–16}. Furthermore, many viral therapeutic treatments and viral vaccines are based on the attenuation of mRNA translation in the viral genes^{17–21}. Down regulation of mRNA translation in an *optimal* manner is also related to fundamental biomedical topics such as molecular evolution and functional genomics^{22–24}.

Here we study for the first time optimal down regulation of translation in a dynamical model of translation. A standard model for translation is the *totally asymmetric simple exclusion process* (TASEP)^{25,26}. In this model, particles hop unidirectionally along an ordered lattice of N sites. Simple exclusion means that a particle cannot hop into a site that is already occupied by another particle. This models hard exclusion between the particles, and creates an indirect coupling between the particles. Indeed, if a particle remains in the same site for a long time then all the particles preceding this site cannot move forward leading to a “traffic jam”. The hops along the lattice

¹School of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. ²Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv 69978, Israel. ³Dept. of Biomedical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel. Correspondence and requests for materials should be addressed to T.T. (email: tamirtul@post.tau.ac.il)

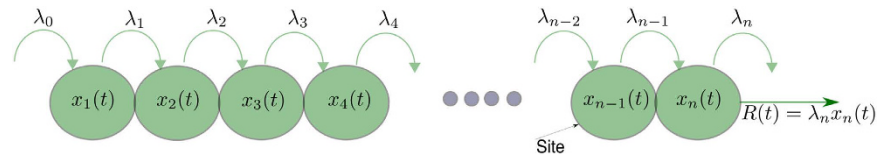


Figure 1. The RFM models unidirectional flow along a chain of n sites. The state variable $x_i(t) \in [0, 1]$ represents the density at site i at time t . The parameter $\lambda_i > 0$ controls the transition rate from site i to site $i + 1$, with $\lambda_0 > 0$ [$\lambda_n > 0$] controlling the initiation [exit] rate. The output rate at time t is $R(t) = \lambda_n x_n(t)$.

are stochastic, and the rate of hopping from site i to site $i + 1$ is denoted by γ_i . A particle can hop to [from] the first [last] site of the lattice at a rate α [β]. The flow through the lattice converges to a steady-state value that depends on N and the vector of parameters

$$\mu := [\alpha, \gamma_1, \dots, \gamma_{N-1}, \beta]'. \quad (1)$$

In the context of translation, the lattice is the mRNA molecule; the particles are the ribosomes; and hard exclusion means that a ribosome cannot move forward if the codon in front of it is covered by another ribosome. In the *homogeneous* TASEP (HTASEP) all the transition rates within the lattice are assumed to be equal and normalized to 1, i.e. $\gamma_i = 1, i = 1, \dots, N - 1$, and thus the model is specified by an input rate α , an exit rate β , and an order N . TASEP is a fundamental model in non-equilibrium statistical mechanics that has been used to model numerous natural and artificial processes including traffic flow, surface growth, communication networks, evacuation dynamics and more^{27,28}.

The *ribosome flow model* (RFM)²⁹ is a nonlinear, continuous-time compartmental model for the unidirectional flow of “material” along a chain of n consecutive compartments (or sites). It can be derived via a mean-field approximation of TASEP^{27,30}. In the RFM, the state variable $x_i(t): \mathbb{R}_+ \rightarrow [0, 1], i = 1, \dots, n$, describes the normalized amount (or density) of “material” in site i at time t , where $x_i(t) = 1$ [$x_i(t) = 0$] indicates that site i is completely full [completely empty] at time t . Thus, the vector $x(t) := [x_1(t) \dots x_n(t)]'$ describes the density profile along the chain at time t . A parameter $\lambda_i > 0, i = 0, \dots, n$, controls the transition rate from site i to site $i + 1$, where λ_0 [λ_n] is the initiation [exit] rate (see Fig. 1). The output rate at time t is $R(t) = \lambda_n x_n(t)$. In the context of translation, the “material” are the moving ribosomes, and each site represents a group of codons, i.e. the mRNA is coarse-grained into n consecutive sites of codons. Thus, $R(t)$, the output flow of ribosomes at time t , is also the *protein production rate* at time t . It is known³¹ that the RFM admits a unique *steady-state* production rate denoted by $R = R(\lambda)$, where $\lambda := [\lambda_0 \dots \lambda_n]'$.

Both TASEP and the RFM are phenomenological models of unidirectional transportation with excluded flow. The RFM approximates the equations describing the stochastic evolution of the site occupancy probabilities in TASEP by ignoring high-order correlations. This provides a good approximation except perhaps at specific parameter values where TASEP undergoes a sharp change (phase transition). The standard TASEP and the RFM also ignore the fact that the moving ribosome actually covers several codons. RFM with extended objects is an important research direction, but not pursued here. However, we show in the Appendix that the RFM, based on coarse graining of the mRNA molecule, provides predictions that agree well with the extended objects TASEP model.

It is important to note that mRNA translation has also been modeled and studied using more detailed stochastic models that include additional features such as both cognate tRNA-capture rates and translocation rates³², competition for near-cognate and non-cognate tRNAs³³, and more. Typically, models with this level of detail can only be studied via simulations. An important advantage of the RFM is that it is amenable to rigorous analysis. This allows deriving general results that hold for any set of feasible parameters values. In order to confirm the validity of the results derived below using the RFM, we compare the results with those obtained using simulations of TASEP. We also include below a detailed example with parameter values derived from the *S. cerevisiae* gene *YBL025W* that encodes the protein *RRN10*.

Here we apply the RFM to analyze how to maximally down-regulate mRNA translation. To do this, we formulate the following general optimization problem. Given an mRNA molecule with n sites, and a convex and compact region of feasible transition rates Ω^{n+1} , find a vector $\lambda^* \in \Omega^{n+1}$ such that $R(\lambda^*) = \min_{\lambda \in \Omega^{n+1}} R(\lambda)$. In other words, the problem is how to select transition rates, within a feasible region, such that the production rate is minimized (see Fig. 2). To the best of our knowledge, this is the first time that such a problem is analyzed in a dynamical model of mRNA translation.

As a concrete example, consider an RFM with dimension n and rates $\bar{\lambda}_0, \dots, \bar{\lambda}_n$. Given a “total reduction budget” $b \in [0, \min\{\bar{\lambda}_i\}]$, define the feasible set $\Omega^{n+1} \subset \mathbb{R}_+^{n+1}$ by

$$\{[\bar{\lambda}_0 - \varepsilon_0 \dots \bar{\lambda}_n - \varepsilon_n]: \varepsilon_i \geq 0, \varepsilon_0 + \dots + \varepsilon_n = b\}.$$

In other words, the feasible set is the set of all the rates obtained by reducing the rates of the given mRNA molecule by the total reduction equal to b . The question is how to distribute the total reduction budget over the rates so as to obtain the minimal possible protein production rate. We prove that:

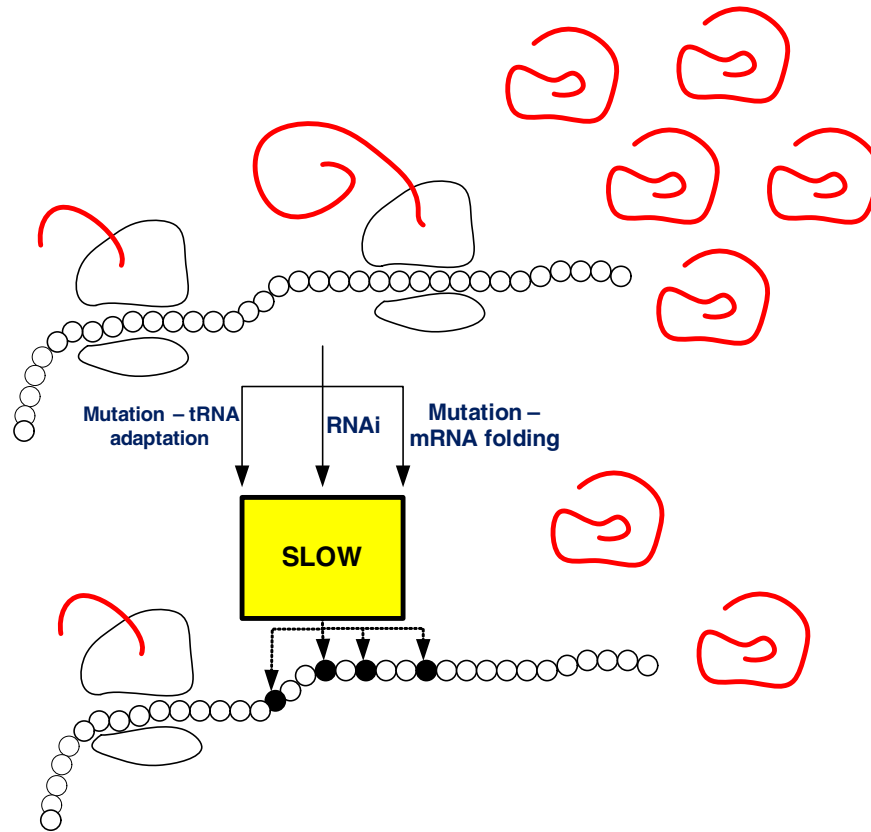


Figure 2. The problem we consider is how to efficiently select transition rates along the mRNA molecule, within a given set of possible rates, such that the protein production rate is minimized. In practice, translation rate modification can be done by introducing mutations into the gene or by designing a corresponding RNAi molecule.

- If some rate $\bar{\lambda}_k$ is a “bottleneck” rate, in a sense that will be made precise below, then an optimal reduction in protein production rate is obtained by using all the reduction budget b to further decrease $\bar{\lambda}_k$;
- If all the given rates are equal, i.e. $\bar{\lambda}_0 = \dots = \bar{\lambda}_n$, then the transition rate at the middle of the mRNA molecule is the bottleneck rate, and thus an optimal reduction in protein production rate is obtained by using all the reduction budget to reduce this transition rate.

Thus, in these cases there exists a single site such that mutating it yields the maximal inhibition of translation. Our results allow to determine where this site is located.

The remainder of this paper is organized as follows. We first briefly review some known results on the RFM that are needed for our purposes. The following section poses the problem of down-regulating the steady-state protein production rate in the RFM in an optimal manner, and then describes our main results. Analysis of the RFM is non-trivial, as this is a nonlinear dynamical model. In particular, the mapping from λ to $R(\lambda)$ is nonlinear and does not admit a closed-form expression. Nevertheless, by combining tools from convex optimization and eigenvalue sensitivity theory, we show that this optimization problem is tractable in some cases, and rigorously prove several results that have interesting biological implications. The final section summarizes and describes several directions for further research. To increase the readability of this paper, all the proofs are placed in Section A in the Appendix. Section B in the Appendix includes a comparison between the RFM and extended objects TASEP predictions showing that the correlation between these two models is very high.

Ribosome Flow Model

The dynamics of the RFM with n sites is given by n nonlinear first-order ordinary differential equations:

$$\begin{aligned}
 \dot{x}_1 &= \lambda_0(1 - x_1) - \lambda_1 x_1(1 - x_2), \\
 \dot{x}_2 &= \lambda_1 x_1(1 - x_2) - \lambda_2 x_2(1 - x_3), \\
 \dot{x}_3 &= \lambda_2 x_2(1 - x_3) - \lambda_3 x_3(1 - x_4), \\
 &\vdots \\
 \dot{x}_{n-1} &= \lambda_{n-2} x_{n-2}(1 - x_{n-1}) - \lambda_{n-1} x_{n-1}(1 - x_n), \\
 \dot{x}_n &= \lambda_{n-1} x_{n-1}(1 - x_n) - \lambda_n x_n.
 \end{aligned} \tag{2}$$

If we define $x_0(t) := 1$ and $x_{n+1}(t) := 0$ then (2) can be written more succinctly as

$$\dot{x}_i = \lambda_{i-1}x_{i-1}(1 - x_i) - \lambda_i x_i(1 - x_{i+1}), \quad i = 1, \dots, n. \tag{3}$$

Eq. (3) can be explained as follows. The change of occupancy at site i at time t is equal to the flow of material from site $i - 1$ to site i minus the flow of material from site i to site $i + 1$ at time t . The latter is $\lambda_i x_i(t)(1 - x_{i+1}(t))$. This flow is proportional to $x_i(t)$, i.e. it increases with the density at site i , and to $(1 - x_{i+1}(t))$, i.e. it decreases as site $i + 1$ becomes fuller. This corresponds to a “soft” version of a simple exclusion principle. Note that the maximal possible flow from site i to site $i + 1$ is the transition rate λ_i .

Let $x(t, a)$ denote the solution of (2) at time $t \geq 0$ for the initial condition $x(0) = a$. Since the state-variables correspond to normalized density levels, with $x_i(t) = 0$ [$x_i(t) = 1$] representing that site i is completely empty [full] at time t , we always assume that a belongs to the closed n -dimensional unit cube: $C^n := \{x \in \mathbb{R}^n: x_i \in [0, 1], i = 1, \dots, n\}$. Let $\text{int}(C^n)$ [∂C^n] denote the interior [boundary] of C^n . It is straightforward to verify that ∂C^n is repelling, i.e. if $a \in \partial C^n$ then $x(t, a) \in \text{int}(C^n)$ for all $t > 0$, so C^n and also $\text{int}(C^n)$ are invariant sets for the dynamics.

An important property of the RFM is the symmetry between the “particles” (i.e. ribosomes) moving from left to right and “holes” (i.e. “lack” of ribosomes) moving from right to left (in the TASEP literature, this property is sometimes referred to as the “particle-hole” symmetry). Indeed, let $q_j(t) := 1 - x_{n+1-j}(t)$, $j = 1, \dots, n$. Then

$$\begin{aligned} \dot{q}_1 &= \lambda_n(1 - q_1) - \lambda_{n-1}q_1(1 - q_2), \\ \dot{q}_2 &= \lambda_{n-1}q_1(1 - q_2) - \lambda_{n-2}q_2(1 - q_3), \\ &\vdots \\ \dot{q}_n &= \lambda_1q_{n-1}(1 - q_n) - \lambda_0q_n. \end{aligned}$$

This is another RFM, but now with rates $\lambda_n, \dots, \lambda_0$.

The RFM has been used to model and analyze the flow of ribosomes along the mRNA molecule during the process of mRNA translation. The (soft) simple exclusion principle corresponds to the fact that ribosomes have volume and cannot overtake one another.

It has been shown in ref. 29 that the correlation between the production rate based on modeling using RFM and using TASEP over all *S. cerevisiae* endogenous genes is 0.96. In addition, it has also been shown there that the RFM agrees well with biological measurements of ribosome densities. Furthermore, it has also been shown that the RFM predictions correlate well (correlations up to 0.6) with protein levels in various organisms (e.g. *E. coli*, *S. pombe*, *S. cerevisiae*). Given the high levels of bias and noise in measurements related to gene expression and the inherent stochasticity of intracellular biological processes (see e.g. refs 34,35), these correlation values demonstrate the relevance of the RFM in this context.

Steady-State Spectral Representation. Ref. 31 has shown that the RFM is a *tridiagonal cooperative dynamical system*³⁶, and that (2) admits a *unique* steady-state point $e = e(\lambda_0, \dots, \lambda_n) \in \text{int}(C^n)$ that is globally asymptotically stable, that is, $\lim_{t \rightarrow \infty} x(t, a) = e$ for all $a \in C^n$ (see also ref. 37). This means that the ribosomal density profile always converges to a steady-state profile that depends on the rates, but not on the initial condition. In particular, the production rate $R(t) = \lambda_n x_n(t)$ converges to a steady-state value $R := \lambda_n e_n$.

At steady-state (i.e., for $x = e$), the left-hand side of all the equations in (2) is zero, so

$$\begin{aligned} \lambda_0(1 - e_1) &= \lambda_1 e_1(1 - e_2) \\ &= \lambda_2 e_2(1 - e_3) \\ &\vdots \\ &= \lambda_{n-1} e_{n-1}(1 - e_n) \\ &= \lambda_n e_n \\ &= R. \end{aligned} \tag{4}$$

This yields

$$R = \lambda_i e_i(1 - e_{i+1}), \quad i = 0, \dots, n, \tag{5}$$

where $e_0 := 1$ and $e_{n+1} := 0$. Ref. 38 used these expressions to provide a *spectral representation* of the mapping from the set of rates λ to the steady-state production rate R . Let $\mathbb{R}_+^n := \{y \in \mathbb{R}^n: y_i \geq 0, i = 1, \dots, n\}$ and $\mathbb{R}_{++}^n := \{y \in \mathbb{R}^n: y_i > 0, i = 1, \dots, n\}$.

Theorem 1 [Ref. 38] *Given an RFM with rates $\lambda = [\lambda_0 \dots \lambda_n]’ \in \mathbb{R}_{++}^{n+1}$, let $R = R(\lambda)$ denote its steady-state production rate. Define an $(n + 2) \times (n + 2)$ Jacobi matrix $A = A(\lambda)$ by*

$$A := \begin{bmatrix} 0 & \lambda_0^{-1/2} & 0 & 0 & \dots & 0 & 0 \\ \lambda_0^{-1/2} & 0 & \lambda_1^{-1/2} & 0 & \dots & 0 & 0 \\ 0 & \lambda_1^{-1/2} & 0 & \lambda_2^{-1/2} & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & \dots & \lambda_{n-1}^{-1/2} & 0 & \lambda_n^{-1/2} \\ 0 & 0 & 0 & \dots & 0 & \lambda_n^{-1/2} & 0 \end{bmatrix}. \tag{6}$$

Then:

1. The eigenvalues of A are real and distinct, and if we order them as $\zeta_1 < \dots < \zeta_{n+2}$ then $\zeta_{n+2} = (R(\lambda))^{-1/2}$.
2. Let $s_i(\lambda) := \frac{\partial}{\partial \lambda_i} R(\lambda)$, i.e. the sensitivity of R with respect to (w.r.t.) the rate λ_i . Let $v \in \mathbb{R}_{++}^{n+2}$ denote an eigenvector of A corresponding to the eigenvalue ζ_{n+2} . Then

$$s_i(\lambda) = \frac{2R^{3/2}}{\lambda_i^{3/2} v^i v} v_{i+1} v_{i+2}, \quad i = 0, \dots, n. \tag{7}$$

This means that the steady-state production rate, and its sensitivity with respect to the transition rates, can be computed efficiently using numerical algorithms for computing the eigenvalues and eigenvectors of tridiagonal matrices. Theorem 1 also implies that

$$R(c\lambda_0, \dots, c\lambda_n) = cR(\lambda_0, \dots, \lambda_n), \quad \text{for all } c > 0, \tag{8}$$

i.e. $R(\lambda)$ is homogeneous of degree one.

Another important implication of Theorem 1 is that R is a *strictly concave* function of the transition rates $\{\lambda_0, \dots, \lambda_n\}$ over \mathbb{R}_{++}^{n+1} ³⁸. Also, it implies that $\frac{\partial}{\partial \lambda_i} R > 0$ for all i , that is, an increase in any of the rates yields an increase in the steady-state production rate.

For more on the analysis of the RFM, and also networks of interconnected RFMs, using tools from systems and control theory, see e.g. refs 38–44.

Main Results

We begin by posing a general minimization problem for the steady-state production rate in the RFM (see Problem 1 below). Using the concavity of the mapping from the rates λ to the steady-state production rate R implies that any solution to this problem must be an extremal point of the feasible set (see Proposition 1 below). We then show how various interesting biological problems can be cast as special cases of this general problem.

Problem 1 Given a convex and compact feasible set of transition rates $\Omega^{n+1} \subset \mathbb{R}_{++}^{n+1}$, find $\lambda^* \in \Omega^{n+1}$ such that $R(\lambda^*) = \min_{\lambda \in \Omega^{n+1}} R(\lambda)$.

From the biological point of view, the feasible set of transition rates Ω^{n+1} depends on all the biophysical constraints on the transition rates along the coding sequence. For example, the maximal/minimal decoding rate of a codon (e.g. via its adaptation to the tRNA pool)⁴⁵, the maximal possible effect of mRNA folding (after codon substitution) on each codon⁴⁶, the maximal possible effect (after amino acid substitution) of the interaction of the ribosome with amino acids of the nascent peptide⁴⁷, and the maximal possible elongation slow down due to interaction with microRNAs^{9,10}.

Below we explain how to pose various interesting biological problems in the framework of Problem 1. Examples include finding the minimal number of mutations that down regulate translation of a gene/mRNA under a certain “total reduction budget”. This is practically important when we use costly (in terms of time and money) gene editing approaches. Another related question is how to down regulate translation of a gene/mRNA with a maximal number of mutations. This is important when attenuating viral replication rate for generating a safe live attenuated vaccine. A large number of mutations reduces the probability of reverting. One may also define the feasible set in Problem 1 in such a way that some rates cannot be changed. This is relevant for example when some codons along the mRNA cannot be modified. Indeed, various positions along the mRNA affect regulatory mechanisms that we may not want to alter (e.g. co-translational folding and splicing).

It is well-known (see, e.g. ref. 48, Thm. 7.42) that if $f: \Omega^{n+1} \rightarrow \mathbb{R}$ is a continuous and strictly convex function defined over a convex and compact set Ω^{n+1} then all the maximizers of f over Ω^{n+1} are extreme points of Ω^{n+1} (for more on the problem of maximizing a convex function, or equivalently, minimizing a concave function, see e.g. ref. 49). Combining this with the fact that R is a strictly concave function of the transition rates over \mathbb{R}_{++}^{n+1} implies the following.

Proposition 1 Every solution of Problem 1 is an extreme point of Ω^{n+1} .

In particular, if the set of extreme points of Ω^{n+1} is finite then one can always solve Problem 1 by simply calculating $R(\lambda)$ for all λ that are extreme points of Ω^{n+1} , and then finding the minimum of these values. In particular, if Ω^{n+1} is a convex polytope then the extreme points are just the vertices of Ω^{n+1} . Thus, when the biophysical constraints lead to a feasible set of rates that is a convex polytope then it is computationally straightforward to

determine how to modify the rates so as to obtain the largest decrease in production rate under reasonable bio-physical constraints.

In the remainder of this section, we consider three special cases of Problem 1 for which it is also possible to obtain analytic results.

Problem 2 Given an RFM with n sites, rates $\bar{\lambda}_0, \dots, \bar{\lambda}_n$, and a total reduction budget $b \in [0, \min\{\bar{\lambda}_i\}]$, let $\Omega^{n+1} = \Omega^{n+1}(\bar{\lambda}, b)$ be the set

$$\left\{ [\bar{\lambda}_0 - \varepsilon_0, \dots, \bar{\lambda}_n - \varepsilon_n]: \varepsilon_i \geq 0, \sum_{i=0}^n \varepsilon_i = b \right\}. \tag{9}$$

Find $\lambda^* \in \Omega^{n+1}$ such that $R(\lambda^*) = \min_{\lambda \in \Omega^{n+1}} R(\lambda)$.

In other words, Ω^{n+1} is the set of all the rates that can be obtained by applying a total reduction b to the given rates $\bar{\lambda}_i$. From a mathematical point of view, b provides a bound on the total possible rate reduction. It also couples the reduction in different rates, as a larger reduction in one rate must be compensated by smaller reductions in the other rates so that the total reduction will not exceed b . From a synthetic biology point of view, b can be used to capture the idea of maximally inhibiting the production rate while minimizing the side-effects of this down regulation. For example, a very small value of b forces a solution with small modifications in all the rates. This is expected of course to minimize the effect of the mutations on the fitness of the cell/organism. Since co-translational folding^{50–52} is related to the ribosome transition rates along the mRNA, smaller changes in the rates are expected to have a smaller effect on protein folding (and thus on the functionality of the protein and the overall organismal fitness). Smaller changes in the transition rates are also related to a “simpler” biological solution in the sense of fewer mutations, less miRNAs, etc.

The next example demonstrates Problem 2.

Example 1 Consider an RFM with length $n = 4$ and transition rates

$$\bar{\lambda}_0 = 0.85, \bar{\lambda}_1 = 0.92, \bar{\lambda}_2 = 0.78, \bar{\lambda}_3 = 0.57, \bar{\lambda}_4 = 0.88.$$

The steady-state production rate is $R(\bar{\lambda}_0, \dots, \bar{\lambda}_4) = 0.2308$ (all numbers are to four digit accuracy). Suppose that the total reduction budget is $b = 0.1$. Then, for example, the vector

$$\lambda := \bar{\lambda} - [0.05 \ 0 \ 0 \ 0.02 \ 0.03]',$$

belongs to Ω^5 , and $R(\lambda) = 0.2260$. An optimal solution for Problem 2 is

$$\lambda^* := [0.85 \ 0.92 \ 0.78 \ 0.47 \ 0.88]' \in \Omega^5,$$

with $R(\lambda^*) = 0.2140$. Note that this corresponds to reducing b from the rate $\bar{\lambda}_3$, which is the minimum of all the rates $\bar{\lambda}_i$, leaving all the other rates unchanged. □

Let $d^i \in \mathbb{R}^{n+1}$ denote the $(i + 1)$ th column of the $(n + 1) \times (n + 1)$ identity matrix. The set $\Omega^{n+1}(\bar{\lambda}, b)$ is a convex polytope with vertices:

$$v^i := [\bar{\lambda}_0 \ \dots \ \bar{\lambda}_n]' - b d^i, \quad i = 0, \dots, n.$$

If there exists an index i such that $\bar{\lambda}_i = b$ then it is clear that an optimal solution is to reduce $\bar{\lambda}_i$ to 0, as then the steady-state production rate will be zero. So we always assume that b takes values in the set $[0, \min\{\bar{\lambda}_i\} - \rho]$, for some $\rho > 0$. This means that Problem 2 is a special case of Problem 1, as $\Omega^{n+1}(\bar{\lambda}, b)$ is a convex polytope contained in \mathbb{R}_{++}^{n+1} .

By Prop. 1, every solution of Problem 2 is contained in the set $\{v^0, \dots, v^n\}$. In other words, every minimizer corresponds to reducing *all* the available budget b from a single rate. This immediately yields a simple and efficient algorithm for solving Problem 2: use the spectral representation of R to compute $R(v^i), i = 0, \dots, n$, and then find the minimum of all these values. Since the matrix A in (6) is symmetric and tridiagonal, calculating $R(v^i)$ can be done efficiently even for large values of n . We wrote a simple (and unoptimized) MATLAB script for solving Problem 2, and ran it on a MAC laptop with a 2.6 GHz Intel core i7 processor. For an RFM with $n = 500$ (a typical coding region includes a few hundred codons⁵³), rates $\bar{\lambda}_i = 1, i = 0, \dots, 500$, and $b = 0.1$, the optimal solution is found in 3.14 seconds.

Example 1 may suggest that reducing the slowest transition rate by b always yields an optimal solution, but in general this is not true (see Example 3 below).

One may also consider a different feasible set in Problem 2, namely,

$$\left\{ [\bar{\lambda}_0 - \varepsilon_0, \dots, \bar{\lambda}_n - \varepsilon_n]: \varepsilon_i \geq 0, \sum_{i=0}^n \varepsilon_i \leq b \right\},$$

i.e. here the total reduction is *up to* b . However, by Theorem 1 $\frac{\partial}{\partial \lambda_i} R(\lambda) > 0$ for all i , and thus an optimal solution for this problem is guaranteed to agree with an optimal solution of Problem 2.

The next example demonstrates the effect of increasing the total reduction rate b on the optimal solution of Problem 2. As noted above, we also compare the results derived for the RFM to TASEP simulations. The simulations used throughout this paper use a parallel update mode. At each time tick t_k , the sites along the lattice are

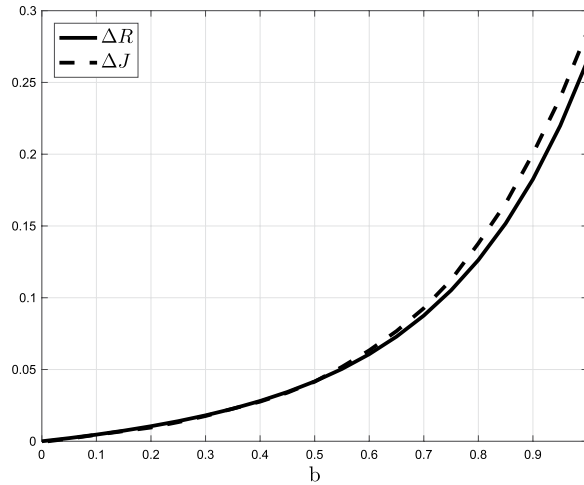


Figure 3. ΔR and ΔJ as a function of b in Example 2.

scanned from site N backwards to site 1. If it is time to hop, and the consecutive site is empty then the particle advances to the consecutive site. If the consecutive site is occupied the next hopping time, $t_k + \varepsilon_k$, is generated randomly. For site i , ε_k is exponentially distributed with parameter $(1/\mu_{i+1})$ (see (1)). The occupancy at each site is averaged throughout the simulation, with the first 10^7 cycles discarded in order to obtain the steady-state value. Let $\rho = \rho(\mu) \in \mathbb{R}_+^N$ denote the averaged occupancy and $J = J(\mu) := \beta \rho_N$ denote the averaged current (or output rate).

Example 2 Consider an RFM with dimension $n = 10$, and rates $\bar{\lambda}_i = 1, i = 0, \dots, n$. Here $R(\bar{\lambda}) = 0.2652$. We calculated the optimal solution λ^* for different values of b , and also the value $\Delta R(b) := R(\bar{\lambda}) - R(\lambda^*)$, that is, the optimal reduction in the protein production rate that can be obtained for various values of b . We also simulated TASEP with $N = 10, \mu = \bar{\lambda}$, and also with the optimal solution $\mu^* := \bar{\lambda}^*$, for different values of b , and computed the value $\Delta J(b) := J(\mu) - J(\mu^*)$. Figure 3 depicts ΔR and ΔJ as a function of b . It may be seen that ΔR and ΔJ increase quickly with b (specifically, the relation is superlinear), and that ΔR provides a good approximation of ΔJ . \square

Optimal reduction and sensitivities. It is also possible to derive theoretical results on the structure of an optimal solution λ^* in Problem 2 using the sensitivities $s_i(\lambda) := \frac{\partial}{\partial \lambda_i} R(\lambda)$ (that can be computed efficiently using (7)). Recall that the proofs of all the analysis results are placed in Section A in the Appendix.

Proposition 2 Consider Problem 2. If there exist $i, j \in \{0, \dots, n\}$ such that

$$s_i(\bar{\lambda}) < s_j(\bar{\lambda}) \tag{10}$$

then any optimal solution λ^* satisfies $\lambda_i^* = \bar{\lambda}_i$.

In other words, if the sensitivity of the steady-state production rate to rate λ_i at $\bar{\lambda}$ is lower than some other sensitivity then an optimal solution will *not* include a reduction in $\bar{\lambda}_i$. Indeed, it is better to distribute the reduction budget over some other, more sensitive, rates.

Remark 1 Note that since R is a strictly concave function of the rates,

$$\frac{\partial}{\partial \lambda_i} s_i(\bar{\lambda}) = \frac{\partial^2}{\partial \lambda_i^2} R(\bar{\lambda}) < 0,$$

for any $\bar{\lambda} \in \mathbb{R}_{++}^{n+1}$ and any $i \in \{0, \dots, n\}$. In other words, a decrease in $\bar{\lambda}_i$ increases the sensitivity w.r.t. this rate.

Proposition 2 leads to the following definition.

Definition 1 Given an RFM with rates $\bar{\lambda}$, a transition rate $\bar{\lambda}_j$ is called a bottleneck rate if $s_j(\bar{\lambda}) > s_i(\bar{\lambda})$, for all $i \neq j$.

In other words, a bottleneck rate is one with a maximal sensitivity.

Combining this with Proposition 2 immediately yields the following result.

Corollary 1 Given an RFM with rates $\bar{\lambda}$, suppose that $s_j(\bar{\lambda})$ is a bottleneck rate. Then the unique optimal solution to Problem 2 is obtained by reducing $\bar{\lambda}_j$ by b .

An important observation is that the slowest rate along the mRNA molecule and the bottleneck rate may be different. The next example demonstrates this.

Example 3 Consider an RFM with dimension $n = 4$, and rates $\bar{\lambda}_3 = 1.85, \bar{\lambda}_i = 2.0, i = 0, 1, 2, 4$. In this case, $s_0(\bar{\lambda}) = 0.0297, s_1(\bar{\lambda}) = 0.0687, s_2(\bar{\lambda}) = 0.0901, s_3(\bar{\lambda}) = 0.0856$, and $s_4(\bar{\lambda}) = 0.0343$. Thus, although the minimal rate is $\bar{\lambda}_3$, the bottleneck rate is $\bar{\lambda}_2$. In particular, the optimal solution will be to reduce $\bar{\lambda}_2$ by b , and not $\bar{\lambda}_3$, even though $\bar{\lambda}_3$ is the minimal rate. \square

However, note that Remark 1 implies that if some rate λ_i is decreased enough then it will eventually become a bottleneck rate.

Proposition 2 can be used to derive analytic results in cases where we can obtain explicit information on the sensitivities at a point $\bar{\lambda} \in \mathbb{R}_+^{n+1}$. The next two results demonstrate this.

Proposition 3 Consider an RFM with dimension n and with equal rates, i.e. $\bar{\lambda}_0 = \dots = \bar{\lambda}_n$. If n is even then the unique optimal solution to Problem 2 is: $\lambda^* = \bar{\lambda} - bd^{n/2}$. If n is odd then there are two optimal solutions: $\lambda^* = \bar{\lambda} - bd^{\lfloor n/2 \rfloor}$ and $\lambda^* = \bar{\lambda} - bd^{\lfloor n/2 \rfloor + 1}$.

In other words, in the case where all the rates are equal, the bottleneck is at the center of the chain. These results are closely related to the fact that in a dynamic model for phosphorelay⁵⁴, that is very similar to the RFM, the middle layer in the model is the most sensitive to changes in the input. This also agrees with the so called “edge-effect” in the HTASEP^{55–57}, i.e. the fact that the steady-state output rate is less sensitive to the rates that are close to the edges of the chain. For more on the sensitivity of TASEP to manipulations in the initiation, hopping, and exit rates, see refs 57–60.

Another case where analytic results can be derived is when the rates in the RFM lead to equal steady-state occupancies along the mRNA molecule. This happens when $\lambda_1 = \lambda_2 = \dots = \lambda_{n-1} = \lambda_0 + \lambda_n$ (see (4)).

Proposition 4 Consider an RFM with dimension n and rates $\bar{\lambda}$ such that $\bar{e}_1 = \dots = \bar{e}_n := e_c$, i.e. all the steady-state occupancies are equal, and e_c denotes their common value.

1. If $e_c < 1/2$ then the unique optimal solution to Problem 2 is

$$\lambda^* = \bar{\lambda} - bd^0. \tag{11}$$

2. If $e_c > 1/2$ then the unique optimal solution to Problem 2 is

$$\lambda^* = \bar{\lambda} - bd^n. \tag{12}$$

3. If $e_c = 1/2$ then (11) and (12) are the optimal solutions.

In other words, if the equal occupancy is relatively low [high] then maximal inhibition of the production rate is obtained by reducing the total reduction rate from the initiation [exit] rate, leaving all the other rates unchanged.

Example 4 Consider Problem 2 for an RFM with $n = 5$, rates $\bar{\lambda} = [1 \ 5/2 \ 5/2 \ 5/2 \ 5/2 \ 3/2]^t$, and $b = 1/2$. Note that in this case $\bar{e}_1 = \dots = \bar{e}_5 = 2/5$. A calculation yields $R(\bar{\lambda} - bd^0) = 0.3999, R(\bar{\lambda} - bd^1) = 0.5651, R(\bar{\lambda} - bd^2) = 0.5762, R(\bar{\lambda} - bd^3) = 0.5829, R(\bar{\lambda} - bd^4) = 0.5874$, and $R(\bar{\lambda} - bd^5) = 0.5746$, so the optimal solution is $\lambda^* = \bar{\lambda} - bd^0$. Since $e_c < 1/2$, this agrees with Proposition 4. We also simulated TASEP with $N = 5, \mu = \frac{2}{5}\bar{\lambda}$ (this scaling is used to make $\mu_i \leq 1$ for all i), and a reduction budget $q := \frac{2}{5}b = 1/5$. The TASEP simulation yields $J(\mu - qd^0) = 0.1595, J(\mu - qd^1) = 0.2302, J(\mu - qd^2) = 0.2324, J(\mu - qd^3) = 0.2337, J(\mu - qd^4) = 0.2344$, and $J(\mu - qd^5) = 0.2221$, so also for TASEP the optimal solution is $\mu^* = \mu - qd^0$. \square

In some cases it may be more natural to define the transition rate reduction in relative rather than absolute terms. This is captured by the following optimization problem.

Problem 3 Given an RFM with n sites, rates $\bar{\lambda}_0, \dots, \bar{\lambda}_n$, and a total reduction budget $q \in [0, 1]$, let $\Gamma^{n+1} = \Gamma^{n+1}(\bar{\lambda}, q) \subset \mathbb{R}_+^{n+1}$ be the set

$$\left\{ [\bar{\lambda}_0(1 - \delta_0), \dots, \bar{\lambda}_n(1 - \delta_n)]: \delta_i \geq 0, \sum_{i=0}^n \delta_i = q \right\}. \tag{13}$$

Find $\lambda^* \in \Gamma^{n+1}$ such that $R(\lambda^*) = \min_{\lambda \in \Gamma^{n+1}} R(\lambda)$.

For $i \in \{0, \dots, n\}$, let $D^i \in \mathbb{R}^{(n+1) \times (n+1)}$ denote the $(n+1) \times (n+1)$ identity matrix, but with entry $(i+1, i+1)$ changed to $1 - q$. The set Γ^{n+1} is a convex polytope with vertices $u^i := D^i \bar{\lambda}, i = 0, \dots, n$. Thus, Problem 3 is also a special case of Problem 1, and so the minimizer λ^* satisfies $\lambda^* \in \{u^0, \dots, u^n\}$.

In practice, each codon (or coding region) admits a minimal and a maximal possible decoding rate. There are also minimal and maximal values for the initiation rate. These bounds are determined by the biophysical properties of the transcript and the intracellular environment. To model this, we can modify the optimization problems described above to include a bound ℓ_i on the maximal allowed reduction of rate i , for $i = 0, \dots, n$. The next problem demonstrates such a modification for Problem 2.

Problem 4 Consider an RFM with n sites and rates $\bar{\lambda}_0, \dots, \bar{\lambda}_n$. Given a total reduction budget $b \in [0, \min\{\bar{\lambda}_i\} - \rho]$, for some $\rho > 0$, and also bounds $0 < \ell_i < \bar{\lambda}_i, i = 0, \dots, n$, with $\sum_{i=0}^n \ell_i > b$, let Ω^{n+1} be as defined in Problem 2, and let

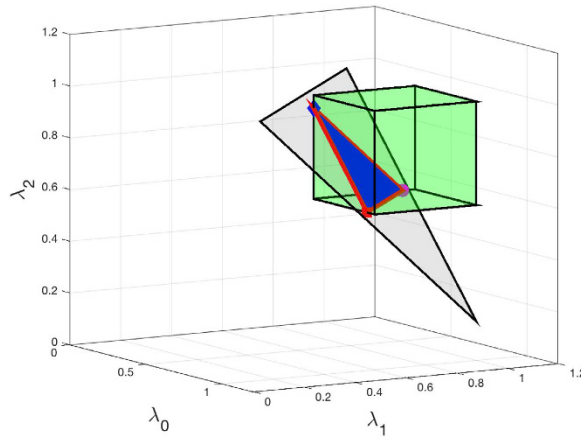


Figure 4. The sets Ω^3 (gray polytope), Ψ^3 (green cube), and Φ^3 (blue polytope) in Example 5.

$$\begin{aligned} \Psi^{n+1} &:= \{\lambda \in \mathbb{R}_{++}^{n+1} : \lambda_i \in [\bar{\lambda}_i - \ell_i, \bar{\lambda}_i], \quad i = 0, \dots, n\}, \\ \Phi^{n+1} &:= \Omega^{n+1} \cap \Psi^{n+1}. \end{aligned} \tag{14}$$

Find $\lambda^* \in \Phi^{n+1}$ such that $R(\lambda^*) = \min_{\lambda \in \Phi^{n+1}} R(\lambda)$.

In other words, the feasible set Φ^{n+1} in Problem 4 is the intersection of the set Ω^{n+1} (defined in Problem 2), and the closed $(n + 1)$ -dimensional cube Ψ^{n+1} that models constraints on the maximal possible reduction of each rate.

Since Φ^{n+1} is compact and convex (being the intersection of two compact and convex sets), Problem 4 admits a solution that is an extreme point of Φ^{n+1} . In general, not all the rates can be reduced by b , and thus an optimal solution may include a reduction of *several* rates.

Example 5 Consider Problem 4 for an RFM with dimension $n = 2$, rates $\bar{\lambda}_i = 1.0$, $i = 0, 1, 2$, and parameters $b = 0.85$, and $\ell_i = 0.4$, $i = 0, 1, 2$. In other words, the total possible reduction is 0.85, but any rate can be reduced by no more than 0.4. Figure 4 depicts the feasible set Φ^3 (blue polytope) that is the intersection of the set Ω^3 (gray polytope) and the set Ψ^3 (green cube). Shown also are the three extreme points of Φ^3 :

$$\begin{aligned} v^1 &:= [0.95 \quad 0.6 \quad 0.6]' \text{ (red circle),} \\ v^2 &:= [0.6 \quad 0.6 \quad 0.95]' \text{ (blue circle),} \\ v^3 &:= [0.6 \quad 0.95 \quad 0.6]' \text{ (magenta circle).} \end{aligned}$$

A calculation yields $R(v^1) = R(v^2) = 0.2538$, whereas $R(v^3) = 0.2764$. It follows that $\lambda^* = v^1$ and $\lambda^* = v^2$ are optimal solutions. Note that these solutions correspond to reducing several rates along the mRNA molecule. Note also that $s(\bar{\lambda}) = [0.1056 \quad 0.1708 \quad 0.1056]'$, so both optimal solutions correspond to a maximal possible reduction in a most sensitive rate, and a maximal possible reduction in another most sensitive rate. \square

In some cases, there may be positions along the coding region that we cannot modify due to their potential effect on various intracellular processes. An important advantage of Problem 4 is that it allows capturing this by simply setting some of the ℓ_i s to zero. If the problem with the additional constraints can be described as a special case of the general Problem 4 then we can efficiently determine the optimal solution.

On the other hand, in down regulation of a viral gene it may be desirable to distribute the synonymous codon modifications over many mRNA sites in order to reduce the chance of spontaneous mutations yielding the original wild type. This is captured by Problem 4 when we set the ℓ_i s to small non-zero values, as then an optimal solution will include a transition rate reduction in many sites.

A biological example. To demonstrate how the results above can be used to analyze translation and provide guidelines for re-engineering the mRNA, we consider the *S. cerevisiae* gene *YBL025W* that encodes the protein *RRN10*, which is related to regulation of RNA polymerase I. This gene has 145 codons (excluding the stop codon). Similarly to the approach used in ref. 29, we divided this mRNA into 6 consecutive pieces: the first piece includes the first 24 codons (that are also related to later stages of initiation⁵²). The other pieces include 25 non-overlapping codons each, except for the last one that includes 21 codons. This partitioning was found to optimize the correlation between the RFM prediction and biological data measurements.

To model this using an RFM with $n = 5$ sites, we first estimated the elongation rates $\lambda_1, \dots, \lambda_5$ using ribo-seq data for the codon decoding rates⁴⁵, normalized so that the median elongation rate of all *S. cerevisiae* mRNAs becomes 6.4 codons per second⁶¹. The site rate is $(\text{site time})^{-1}$, where site time is the sum over the decoding times of all the codons in this site. These rates thus depend on various factors including availability of tRNA molecules,

amino acids, Aminoacyl tRNA synthetase activity and concentration, and local mRNA folding^{1,45,52}. Note that if we replace a codon in a site of mRNA by a synonymous slower codon then the decoding time increases and thus the rate associated with this site decreases.

The initiation rate (that corresponds to the first piece) was estimated based on the ribosome density per mRNA levels, as this value is expected to be approximately proportional to the initiation rate when initiation is rate limiting^{29,62}. Again we applied a normalization that brings the median initiation rate of all *S. cerevisiae* mRNAs to be 0.8⁶³. Adding the initiation time (1/0.4482) to the site time of the first piece yields an RFM with $n = 5$ and parameters:

$$[\bar{\lambda}_0 \dots \bar{\lambda}_5] = [0.1678 \ 0.2572 \ 0.2758 \ 0.2514 \ 0.2612 \ 0.3002].$$

A calculation yields that the steady-state production rate in this RFM is $R = 0.0732$. We also simulated TASEP with $N = 5$ and $\mu = \bar{\lambda}$, and for these values the simulation yields the steady-state current $J = 0.0776$.

In order to analyze the solution of Problem 2 for this RFM we calculated the sensitivities using (7). This yields: $s(\bar{\lambda}) = [0.0795 \ 0.0669 \ 0.0611 \ 0.0578 \ 0.0328 \ 0.0092]$, so $\bar{\lambda}_0$ is a bottleneck rate. This means that the solution for Problem 2 is to reduce all the reduction budget b from $\bar{\lambda}_0$. In biological terms, this suggests that maximal inhibition of production should be based on replacing some (or all) of the first 24 codons with slower synonymous codons. For comparison with the optimization scenarios described below, consider the total budget $b = 0.0089$. The solution for Problem 2 is then to reduce λ_0 by b , and this yields

$$R^* = 0.0725. \quad (15)$$

Reducing λ_0 by b in the model is possible by substituting codons in the first site with their slowest synonymous mutation (for example, the third codon AGA should be replaced by the synonymous codon CGG, increasing the codon decoding time from 0.1128 seconds to 0.2246 seconds). A TASEP simulation with α reduced by b yields

$$J^* = 0.0767. \quad (16)$$

Now suppose that we are not interested in modifying these codons because in this region there are various regulatory signals that we may not want to change (see, for example, ref. 52). To maximize inhibition of production rate under this constraint, we apply Problem 4, with $\ell_0 = 0$, and $\ell_i > b$ for all $i \neq 0$. Now the optimal solution is to reduce b from $\bar{\lambda}_1$. Note that $\bar{\lambda}_1$ has the second largest sensitivity. This yields $R^* = 0.0726$, and is, as expected, higher than the value in (15). Again, the biological data shows that such a reduction can be done by synonymously replacing codons 34 (GCT with GCA), 35 (GTT with GTA), 36 (CCT with CCC), 38 (CCG with CCC), 39 (TTC with TTT), and 49 (GTG with GTA). A TASEP simulation with γ_1 reduced by b yields $J = 0.0769$, which is indeed higher than the value in (16).

Finally, to demonstrate mutations in multiple sites, we used the data to find a scenario where a set of mutations yields the same total decrease in the rates. This can be done by synonymously replacing codons 21 (GTG with GTA), 29 (GAA with GAG), 58 (TTC with TTT), 82 (AAG with AAA), 110 (CTA with CTG), and 141 (GCG with GCA), leading to

$$\lambda = [0.1677 \ 0.2557 \ 0.2733 \ 0.2489 \ 0.2599 \ 0.2991]'$$

Note that all the rates are reduced and that the total reduction is b . This yields $R = 0.0727$, which is again higher than the value in (15). A TASEP simulation with $\mu = \lambda$ yields $J = 0.0771$, which is again higher than the value in (16).

Discussion

There are several approaches for effectively down-regulating translation. Global down-regulation can be achieved by controlling basic translation factors or by using drugs that induce ribosome stalling^{64–66}.

Here we consider down regulation of specific genes via targeting specific codons/regions in these genes. This leads to the problem of finding the codon regions that have the most effect on the steady-state production rate. We pose this problem of optimal down regulation of mRNA translation as a general optimization problem (Problem 1 above) for a mathematical model for ribosome flow, the RFM. All possible modifications of the rates define a feasible set of rates, and, under certain conditions, we give a simple algorithm for finding the optimal solution, that is, the rates that lead to a maximal decrease in the protein production rate. For some specific cases, we also derive theoretical results on the optimal solution. Our general formulation thus provides guidelines on how to pose the optimal down regulation problem properly.

Our results show that the solution must focus on the positions along the mRNA molecule where the transition rate has the strongest effect on the protein production rate. However, this position is not necessarily the one with the minimal rate (though in many cases there are correlations between the two definitions). Many previous studies in the field emphasized the importance of the translation bottleneck^{22,59,67}, however, this is always defined as the minimal rate. We believe that the sensitivity of the coding region sites should be further studied in order to understand better the evolution of transcripts and their design.

In addition, we demonstrated using TASEP simulations that the analytical results obtained for the RFM may be used to obtain an optimal solution in the nonhomogeneous TASEP as well. This is important since TASEP and its many variants are employed to analyze many biological (and other) phenomena, and the analytical results obtained in this paper may be useful in these cases as well.

The optimization problems posed here are flexible enough to capture various scenarios. For example, in some cases it may be desirable to introduce a *minimal* number of changes in the transcript to obtain the desired decrease

in the translation rate. Indeed, generating mutations and using suitable RNAi molecules is costly in time and money. Also, any change in the translation rates can affect various important phenomena such as co-translational folding^{50–52}, as well as other properties that are encoded in the coding region^{52,68,69}. In other cases, such as generating a down-regulated virus strain, it may be desirable to introduce as many mutations as possible.

There are various approaches for synthesizing molecules that block mRNA translation (see e.g. http://www.gene-tools.com/choosing_the_optimal_target). In practice, when determining an optimal position to target (e.g. with RNAi molecules) one must take into account additional biophysical aspects. For example, the GC content at the different regions along the mRNA, the folding of the mRNA, the potential binding affinity of the RNAi and the mRNA, potential un-desired binding of the RNAi to additional mRNAs or regions within the mRNA, etc. Nevertheless, we feel that our results can be integrated to improve the design of such tools.

In addition, it is entirely possible that in the future this downregulation problem will be studied using other and perhaps more detailed models of translation. For example, in practice, there are many mRNA molecules in the cell and they all compete for the finite pool of free ribosomes. In particular, if more ribosomes are stuck in a traffic jam on a certain mRNA molecule then the pool of free ribosomes is depleted yielding a reduction in the production rates in other mRNA molecules. The RFM is a model for ribosome flow along a single isolated mRNA molecule. This is a reasonable model when the expression levels (e.g. the mRNA levels and the total number of ribosomes on the mRNA molecules related to the gene) are relatively low, so that changes in the translation dynamics on one mRNA have a negligible effect on the pool of ribosomes and thus on the other mRNAs. A model for a network of RFMs, interconnected via a dynamic pool of free ribosomes, has been studied in ref. 44. It may be of interest to study the problem of down regulation of a specific mRNA molecule within this framework. In this case, one can also down regulate the mRNA indirectly by affecting the ribosomal pool. However, the analytic tools used here do not directly apply, as the convexity results for a single chain do not necessarily carry over to the case of a network of RFMs.

The results here suggest several biological experiments for studying the problem of optimal down regulation and, in particular, validating the theoretical predictions derived using the RFM. Libraries encoding the same protein using mRNAs with different codons (but similar mRNA levels and translation initiation rates) can be generated as was done in ref. 17. For each variant the protein levels, that are expected to monotonically increase with the production rate²⁹, can be measured either via a reporter protein¹⁷ or directly⁷⁰. The codon decoding rates can be estimated based on ribo-seq experiments^{17,45}. Such an experimental testbed can be used to validate the results reported in this study.

References

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts & P. Walter. *Molecular Biology of the Cell*. (New York: Garland Science, 2002).
2. Y. Arava, Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown & D. Herschlag. “Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*”, *Proceedings of the National Academy of Sciences* **100**(7), 3889–3894 (2003).
3. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman & J. S. Weissman. “Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling”, *Science* **324**(5924), 218–223 (2009).
4. X. Yan, T. A. Hoek, R. D. Vale & M. E. Tanenbaum. “Dynamics of translation of single mRNA molecules *in vivo*”, *Cell* **165**(4), 976–89 (2016).
5. B. Wu, C. Eliscovich, Y. Yoon & R. Singer. “Translation dynamics of single mRNAs in live cells and neurons”, *Science* **352**(6292), 1430–5 (2016).
6. T. Morisaki, K. Lyon, K. F. DeLuca, J. G. DeLuca, B. P. English, Z. Zhang, L. D. Lavis, J. B. Grimm, S. Viswanathan, L. L. Looger, T. Lionnet & T. J. Stasevich. “Real-time quantification of single RNA translation dynamics in living cells”, *Science* **352**(6292), 1425–9 (2016).
7. C. Wang, B. Han, R. Zhou & X. Zhuang. “Real-time imaging of translation on single mRNA transcripts in live cells”, *Cell* **165**(4), 990–1001 (2016).
8. N. Stern-Ginossar & N. T. Ingolia. “Ribosome profiling as a tool to decipher viral complexity”, *Annu. Rev. Virol.* **2**, 335–49 (2015).
9. M. Ghildiyal & P. Zamore. “Small silencing RNAs: an expanding universe”, *Nature Rev. Genet.* **10**, 94–108 (2009).
10. M. Inui, G. Martello & S. Piccolo. “MicroRNA control of signal transduction”, *Nat Rev Mol Cell Biol.* **11**(4), 252–263 (2010).
11. M. Fabian, N. Sonenberg & W. Filipowicz. “Regulation of mRNA translation and stability by microRNAs”, *Annu. Rev. Biochem.* **79**, 351–79 (2010).
12. W. Filipowicz, S. Bhattacharyya & N. Sonenberg. “Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight?” *Nat. Rev. Genet.* **9**(2), 102–14 (2008).
13. C. Croce. “Causes and consequences of microRNA dysregulation in cancer”, *Nat. Rev. Genet.* **10**(10), 704–14 (2009).
14. S. Tavazoie, C. Alarcón, T. Oskarsson, D. Padua, Q. Wang, P. Bos & W. G. J. Massagué. “Endogenous human microRNAs that suppress breast cancer metastasis”, *Nature* **451**(7175), 147–52 (2008).
15. L. Zhang, N. Yang, A. Mohamed-Hadley, S. Rubin & G. Coukos. “Vector-based RNAi, a novel tool for isoform-specific knock-down of VEGF and anti-angiogenesis gene therapy of cancer”, *Biochem. Biophys. Res. Commun.* **303**(4), 1169–78 (2003).
16. G. Devi. “siRNA-based approaches in cancer therapy”, *Cancer Gene Ther.* **13**(9), 819–29 (2006).
17. T. Ben-Yehzkel, S. Atar, H. Zur, A. Diamant, E. Goz, T. Marx, R. Cohen, A. Dana, A. Feldman, E. Shapiro & T. Tuller. “Rationally designed, heterologous *S. cerevisiae* transcripts expose novel expression determinants”, *RNA Biol.* **12**, 972–84 (2015).
18. E. Goz & T. Tuller. “Widespread signatures of local mRNA folding structure selection in four Dengue virus serotypes”, *BMC Genomics* **16** (Suppl 10:S4) (2015).
19. Q. Wang, C. Contag, H. Ilves, B. Johnston & R. Kaspar. “Small hairpin RNAs efficiently inhibit hepatitis C IRES-mediated gene expression in human tissue culture cells and a mouse model”, *Molecular Therapy* **12**(3), 562–8 (2005).
20. J. Coleman, D. Papamichail, S. Skiena, B. Futcher, E. Wimmer & S. Mueller. “Virus attenuation by genome-scale changes in codon pair bias”, *Science* **320**, 1784–7 (2008).
21. J. Perez, A. Pham, M. Lorini, M. Chua, J. Steel & B. tenOever. “MicroRNA-mediated species-specific attenuation of influenza A virus”, *Nat. Biotechnol.* **27**(6), 572–6 (2009).
22. T. Tuller, A. Carmi, K. Vestsigian, S. Navon, Y. Dorfan, J. Zaborske, T. Pan, O. Dahan, I. Furman & Y. Pilpel. “An evolutionarily conserved mechanism for controlling the efficiency of protein translation”, *Cell* **141**(2), 344–54 (2010).
23. J. Forman, A. Legesse-Miller & H. Collier. “A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence”, *Proceedings of the National Academy of Sciences* **105**(39), 14 879–84 (2008).
24. P. Fang, C. Spevak, C. Wu & M. Sachs. “A nascent polypeptide domain that can regulate translation elongation”, *Proceedings of the National Academy of Sciences* **101**(12), 4059–64 (2004).

25. L. B. Shaw, R. K. P. Zia & K. H. Lee. “Totally asymmetric exclusion process with extended objects: a model for protein synthesis”, *Phys. Rev. E* **68**, 021910 (2003).
26. R. Zia, J. Dong & B. Schmittmann. “Modeling translation in protein synthesis with TASEP: A tutorial and recent developments”, *J. Statistical Physics* **144**, 405–428 (2011).
27. A. Schadschneider, D. Chowdhury & K. Nishinari. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. (Elsevier, 2011).
28. S. Srinivasa & M. Haenggi. “A statistical mechanics-based framework to analyze ad hoc networks with random access”, *IEEE Trans. Mobile Computing* **11**, 618–630 (2012).
29. S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppín & T. Tuller. “Genome-scale analysis of translation elongation with a ribosome flow model”, *PLoS Computational Biology* **7**, e1002127 (2011).
30. R. A. Blythe & M. R. Evans. “Nonequilibrium steady states of matrix-product form: a solver’s guide”, *J. Phys. A: Math. Theor.* **40**(46), R333–R441 (2007).
31. M. Margaliot & T. Tuller. “Stability analysis of the ribosome flow model”, *IEEE/ACM Trans. Computational Biology and Bioinformatics* **9**, 1545–1552 (2012).
32. L. Ciandrini, I. Stansfield & M. C. Romano. “Ribosome traffic on mRNAs maps to gene ontology: Genome-wide quantification of translation initiation rates and polysome size regulation”, *PLoS Comput. Biol.* **9**, 1–10 (2013).
33. K. B. Gromadski & M. V. Rodnina. “Kinetic determinants of high-fidelity tRNA discrimination on the ribosome”, *Molecular Cell* **13**, 191–200 (2004).
34. A. Diamant & T. Tuller. “Estimation of ribosome profiling performance and reproducibility at various levels of resolution”, *Biol Direct.* **11**, 24 (2016).
35. M. Kaern, T. C. Elston, W. J. Blake & J. J. Collins. “Stochasticity in gene expression: from theories to phenotypes”, *Nat. Rev. Genet.* **6**, 451–64 (2005).
36. H. L. Smith. *Monotone Dynamical Systems: An Introduction to the Theory of Competitive and Cooperative Systems*, ser. Mathematical Surveys and Monographs. Providence, RI: Amer. Math. Soc. **41** (1995).
37. M. Margaliot, E. D. Sontag & T. Tuller. “Entrainment to periodic initiation and transition rates in a computational model for gene translation”, *PLoS One* **9**(5), e96039 (2014).
38. G. Póker, Y. Zarai, M. Margaliot & T. Tuller. “Maximizing protein translation rate in the nonhomogeneous ribosome flow model: A convex optimization approach”, *J. Royal Society Interface* **11**(100), 20140713 (2014).
39. Y. Zarai, M. Margaliot & T. Tuller. “Explicit expression for the steady-state translation rate in the infinite-dimensional homogeneous ribosome flow model”, *IEEE/ACM Trans. Computational Biology and Bioinformatics* **10**, 1322–1328 (2013).
40. G. Póker, M. Margaliot & T. Tuller. “Sensitivity of mRNA translation”, *Sci. Rep.* **5**(12795) (2015).
41. M. Margaliot & T. Tuller. “Ribosome flow model with positive feedback”, *J. Royal Society Interface* **10**, 20130267 (2013).
42. A. Raveh, Y. Zarai, M. Margaliot & T. Tuller. “Ribosome flow model on a ring”, *IEEE/ACM Trans. Computational Biology and Bioinformatics* **12**(6), 1429–1439 (2015).
43. Y. Zarai, M. Margaliot, E. D. Sontag & T. Tuller. “Controlling mRNA translation”, Submitted. [Online]. Available: <http://arxiv.org/abs/1602.02308> (2016).
44. A. Raveh, M. Margaliot, E. D. Sontag & T. Tuller. “A model for competition for ribosomes in the cell”, *J. Royal Society Interface* **13**(116) (2016).
45. A. Dana & T. Tuller. “Mean of the typical decoding rates: a new translation efficiency index based on ribosome analysis data”, *G3: Genes, Genomes, Genetics* (2014).
46. T. Tuller, I. Vekslar, N. Gazit, M. Kupiec, E. Ruppín & M. Ziv. “Composite effects of the coding sequences determinants on the speed and density of ribosomes”, *Genome Biol.* **12**(11), R110 (2011).
47. R. Sabi & T. Tuller. “A comparative genomics study on the effect of individual amino acids on ribosome stalling”, *BMC Genomics* (2015).
48. A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. Philadelphia, PA: Society for Industrial and Applied Mathematics (2014).
49. R. Enghat. “An algorithm for maximizing a convex function over a simple set”, *J. Global Optimization* **8**(4), 379–391 (1996).
50. G. Zhang, M. Hubalewska & Z. Ignatova. “Transient ribosomal attenuation coordinates protein synthesis and co-translational folding”, *Nat. Struct. Mol. Biol.* **16**(3), 274–80 (2009).
51. S. Pechmann & J. Frydman. “Evolutionary conservation of codon optimality reveals hidden signatures of cotranslational folding”, *Nat. Struct. Mol. Biol.* **20**(2), 237–43 (2013).
52. T. Tuller & H. Zur. “Multiple roles of the coding sequence 5’ end in gene expression regulation”, *Nucleic Acids Res.* **43**(1), 13–28 (2015).
53. J. Zhang. “Protein-length distributions for the three domains of life”, *Trends Genet.* **16**(3), 107–9 (2000).
54. A. Csikasz-Nagy, L. Cardelli & O. S. Soyer. “Response dynamics of phosphorelays suggest their potential utility in cell signaling”, *J. Royal Society Interface* **8**, 480–488 (2011).
55. J. J. Dong, B. Schmittmann & R. K. P. Zia. “Towards a model for protein production rates”, *J. Statistical Physics* **128**(1–2), 21–34 (2007).
56. J. J. Dong, R. K. P. Zia & B. Schmittmann. “Understanding the edge effect in TASEP with mean-field theoretic approaches”, *J. Phys. A: Math. Gen.* **42**(1), 015002 (2009).
57. J. J. Dong, B. Schmittmann & R. K. P. Zia. “Inhomogeneous exclusion processes with extended objects: The effect of defect locations”, *Phys. Rev. E* **76**, 051113 (2007).
58. M. E. Fouladivand, A. B. Kolomeisky & H. Teymouri. “Asymmetric exclusion processes with disorder: Effect of correlations”, *Physical Review E* **78**(6), 061116 (2008).
59. T. Chou & G. Lakatos. “Clustered bottlenecks in mRNA translation and protein synthesis”, *Phys. Rev. Lett.* **93**, 198101 (2004).
60. G. Tripathy & M. Barma. “Driven lattice gases with quenched disorder: exact results and different macroscopic regimes”, *Phys. Rev. E* **58**, 1911–1926 (1998).
61. T. V. Karpinet, D. J. Greenwood, C. E. Sams & J. T. Ammons. “RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis”, *BMC Biol.* **4**(30), 274–80 (2006).
62. M. Margaliot & T. Tuller. “On the steady-state distribution in the homogeneous ribosome flow model”, *IEEE/ACM Trans. Computational Biology and Bioinformatics* **9**, 1724–1736 (2012).
63. D. Chu, E. Kazana, N. Bellanger, T. Singh, M. F. Tuite & T. von der Haar. “Translation elongation can control translation initiation on eukaryotic mRNAs”, *EMBO J.* **33**(1), 21–34 (2014).
64. M. Greenberg, A. Hermanowski & E. Ziff. “Effect of protein synthesis inhibitors on growth factor activation of c-fos, c-myc, and actin gene transcription”, *Mol. Cell Biol.* **6**(4), 1050–7 (1986).
65. M. Clemens, M. Bushell, I. Jeffrey, V. Pain & S. Morley. “Translation initiation factor modifications and the regulation of protein synthesis in apoptotic cells”, *Cell Death Differ.* **7**(7), 603–15 (2000).
66. M. Kozak. “Regulation of translation in eukaryotic systems”, *Annu. Rev. Cell Biol.* **8**, 197–225 (1992).
67. S. Zhang, E. Goldman & G. Zubay. “Clustering of low usage codons and ribosome movement”, *J. Theor. Biol.* **170**, 339–54 (1994).

68. L. Cartegni, S. Chew & A. Krainer. "Listening to silence and understanding nonsense: exonic mutations that affect splicing", *Nat. Rev. Genet.* **3**, 285–98 (2002).
69. A. Stergachis, E. Haugen, A. Shafer, W. Fu, B. Vernot, A. Reynolds, A. Raubitschek, S. Ziegler, E. LeProust, J. Akey & J. Stamatoyannopoulos. "Exonic transcription factor binding directs codon choice and affects protein evolution", *Science* **342**, 1367–72 (2013).
70. B. Schwanhausser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen & M. Selbach. "Global quantification of mammalian gene expression control", *Nature* **473(7347)**, 1367–72 (2011).

Author Contributions

Y.Z., M.M. and T.T. performed the research and wrote the paper.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zarai, Y. *et al.* Optimal Down Regulation of mRNA Translation. *Sci. Rep.* **7**, 41243; doi: 10.1038/srep41243 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017