Thermodynamically-consistent, reduced models of gene regulatory networks

Michael Pan[1], Peter J. Gawthrop[2], Matthew Faria[2], Stuart T. Johnston[1*]

[1]School of Mathematics and Statistics, The University of Melbourne, Parkville, 3010, Victoria, Australia.
[2]Department of Biomedical Engineering, The University of Melbourne, Parkville, 3010, Victoria, Australia.

*Corresponding author(s). E-mail(s): stuart.johnston@unimelb.edu.au; Contributing authors: pan.m@unimelb.edu.au; peter.gawthrop@unimelb.edu.au; matthew.faria@unimelb.edu.au;

Abstract

Synthetic biology aims to engineer novel functionalities into biological systems. While the approach has been widely applied to single cells, the scale of synthetic circuits designed in this way is limited by factors such as resource competition and retroactivity. Synthetic biology of cell populations has the potential to overcome some of these limitations by physically isolating synthetic genes from each other. To rationally design cell populations, we require mathematical models that link between intracellular biochemistry and intercellular interactions. The interfacing of agent-based models with systems biology models is particularly important in understanding the effects of cell heterogeneity and cell-to-cell interactions.

In this study, we develop a model of gene expression that is suitable for incorporation into agent-based models of cell populations. To be scalable to large cell populations, models of gene expression should be both computationally efficient and compliant with the laws of physics. We satisfy the first requirement by applying a model reduction scheme to translation, and the second requirement by formulating models using bond graphs. We show that the reduced model of translation faithfully reproduces the behaviour of the full model at steady state. The reduced model is benchmarked against the full model, and we find a substantial speedup at realistic protein lengths. Using the modularity of the bond graph approach, we couple separate models of gene expression to build models of the toggle switch and repressilator. With these models, we explore the effects of

resource availability and cell-to-cell heterogeneity on circuit behaviour. The modelling approaches developed in this study are a bridge towards rationally designing collective cell behaviours such as synchronisation and division of labour.

Keywords: synthetic biology, bond graph, gene regulatory networks, thermodynamics

1 Introduction

Synthetic biology is the engineering of novel functionalities into biological systems. This enables biological solutions to problems that are difficult or impossible to solve with conventional engineering technologies, such as the detection and breakdown of toxic environmental pollutants and the production of biological therapeutics (Gray et al, 2018; Church et al, 2014). Synthetic biology has been applied with great success to single cells, making use of synthetic genes as the fundamental building block. The field first gained attention through early biological implementations of two simple genetic circuits; the genetic *toggle switch*, which is capable of flipping between two stable states (Gardner et al, 2000); and the *repressilator*, a negative feedback loop composed of three repressor genes capable of oscillations (Elowitz and Leibler, 2000). Since then, synthetic biology has been used to design more complex systems such as biosensors and cell computers (Shaw et al, 2019; Gander et al, 2017). To scale up from simple to complex systems, synthetic biology employs a modular approach from engineering. Thus, standard modules (genetic constructs that have been designed according to specifications) can be re-used as part of a more complex system. With this approach, even relatively simple genetic circuits can be assembled into biological logic gates, and ultimately biological devices capable of making complex decisions (Gander et al, 2017).

There are several known limitations to single-cell synthetic biology. Firstly, there may be undesired cross-talk between synthetic genes, leading to unpredictable function (Slusarczyk et al, 2012). Secondly, the newly inserted genes consume the same resources (energy, ribosomes and amino acids) as other pathways in the cell (Boo et al, 2019). Thus, a resource-intensive gene may compete with the cell's endogenous pathways, leading to poor growth (Gorochowski, 2016). Finally, it cannot be assumed that upstream modules are completely isolated from the effects of downstream modules. In general, downstream pathways will influence and in some cases disrupt the behaviour of upstream pathways. This is known as retroactivity, and can lead to poor modularity even in the absence of cross-talk (Del Vecchio, 2013).

To address the known limitations of single-cell synthetic biology, synthetic biology of cell populations and consortia has been proposed (Gorochowski, 2016; Caringella et al, 2023). Synthetic modules can be inserted into different cells of a consortium, physically isolating them from each other (Shaw et al, 2019). This reduces the risk of cross-talk, competition and retroactivity between modules. Furthermore, it is possible to design population-level behaviours that would not be possible within homogeneous populations. For example, the quorum sensing system in bacteria is a frequent target of

synthetic biology (Boo et al, 2021). This system senses local cell density via the accumulation of signalling molecules, and can be engineered to reduce noise and synchronise oscillations (Gorochowski, 2016). It is also possible to engineer cell populations to form various spatial patterns when grown on media (Basu et al, 2005).

In spite of the clear potential of cell populations for synthetic biology, it is still an emerging field compared to single-cell synthetic biology. A major challenge is predicting how changes at the single-cell level affect behaviour at the population scale. Because it is infeasible to physically test every possible implementation of a synthetic biological system, synthetic biologists use design-build-test-learn cycles to optimise performance with limited time and resources (Gray et al, 2018). In this context, mathematical models are essential to ruling out infeasible designs prior to implementation. Mathematical models have been widely used to describe cellular biochemistry (systems biology) and interactions between cells (agent-based models) (Machado et al, 2011). While it is possible to model cell populations without agent-based models, coupling systems biology models to agent-based models is essential to understanding phenomena such as cell heterogeneity, synchronisation and cell migration. Such population-level dynamics pose a unique set of challenges for modelling engineered cell populations. In particular, cell populations can easily consist of millions of cells (Prindle et al, 2012). Thus, while complex models of gene regulatory networks are sufficient for single-cell synthetic biology, computational efficiency becomes a major consideration in engineering cell populations (Nagarajan et al, 2022).

Because synthetic biology is focussed on engineering gene circuits, modelling gene regulatory networks is a key point of interest. The landmark studies implementing the toggle switch and repressilator used simple ordinary differential equation models of gene regulation, ignoring the dynamics of transcription and translation (Gardner et al, 2000; Elowitz and Leibler, 2000). This approach has been generalised and expanded upon to model large-scale gene regulatory networks (Ryall et al, 2012). However, the above models do not explicitly model transcription and translation, nor do they account for ribosome and energy availability. More recent models have sought to incorporate this level of detail. Whole-cell models of *Mycoplasma genitalium* and *Escherichia coli* account for the use of ribosomes, amino acids and energy in translation (Karr et al, 2012; Macklin et al, 2020). Weiße et al (2015) take a simpler approach, using generic molecules for energy and monomers as variables in a differential equation model of transcription and translation. There are also more detailed models of translation that account for the position of ribosomes on mRNA and can predict the occurrence of ribosomal 'traffic jams'. These include the stochastic totally asymmetric exclusion process (TASEP) and its deterministic approximation the ribosome flow model (RFM) (Raveh et al, 2016). However, a consideration for the thermodynamics of gene expression has for the most part been lacking. Many models completely ignore the energy molecule adenosine triphosphate (ATP) and therefore do not account for the dependence of expression dynamics on energy. This is surprising given that translation consumes four ATP molecules per amino acid (Kempes et al, 2017), making it one of the most energy-consuming processes in a cell (30-75%) (Karr et al, 2012; Shallom et al, 2021). If we are to adequately model gene regulatory networks for the synthetic biology of cell populations, a new modelling approach is needed that accounts

for energy while also resulting in models simple enough to be incorporated into large cell populations.

In this study, we propose a new, simplified, and thermodynamically consistent model of gene expression. The model satisfies three key criteria: it is (i) computationally feasible to simulate; (ii) dependent on the availability of energy and ribosomes; and (iii) thermodynamically consistent. We primarily focus on translation as it is a major consumer of energy. A model reduction scheme is developed to simplify the chains of reactions involved in translation elongation. This ensures that the model is simple but biochemically plausible, i.e. it is a limiting case of a more complex reaction network. To ensure thermodynamic consistency, we use the bond graph approach, which is a powerful, modular framework for modelling biological systems in a physically consistent manner (Gawthrop and Pan, 2022). Previous work has shown that thermodynamic constraints automatically fall out from the parameterisation inherent to bond graph models, and that the approach is scalable to large biochemical networks (Gawthrop and Crampin, 2014; Pan et al, 2018, 2021). Furthermore, bond graphs provide a natural methodology for accounting for retroactivity (Gawthrop, 2021). We exploit the graphical, modular nature of bond graphs to couple models of gene expression together, enabling the construction of common synthetic biology motifs such as the toggle switch and repressilator.

In Sect. 2, we introduce the bond graph formalism, which is then used to construct thermodynamically consistent models of gene regulatory networks. A model reduction scheme for translation is developed to ensure the model is simple but still retains important features of a more detailed model (Sect. 2.4). By coupling models of gene expression together, models of synthetic circuits are developed (Sect. 2.6). We first compare the full and reduced models, and show that their steady-state behaviours match well (Sect. 3.1). We next show that the bond graph model is able to reproduce experimentally observed behaviours in common synthetic biology motifs. In Sects. 3.3–3.4, we investigate the effects of energy and ribosome availability on synthetic circuits. We then examine the potential effects of parameter heterogeneity in synthetic circuits (Sect. 3.5). Finally, Sect. 4 contains a discussion of the results and concluding remarks.

2 Methods

2.1 Overview

The gene is a key building block of synthetic biology. Here, we consider two key processes in gene expression: *transcription*, which produces mRNA from a DNA template; and *translation*, which produces protein from an mRNA template (Fig. 1). This section will detail the methods required to model gene expression in a thermodynamic framework. Inspired by Weiße et al (2015), we adopt a simplistic approach for modelling these processes. In particular, we use generic species to model cell resources rather than accounting for all their constituent biomolecules. Thus, both energy (A) and ribosomes (R) are represented as individual species. Fig. 1 depicts an overview of the model. It is assumed that the reactions take place in a constant volume, i.e. in the absence of cell expansion.

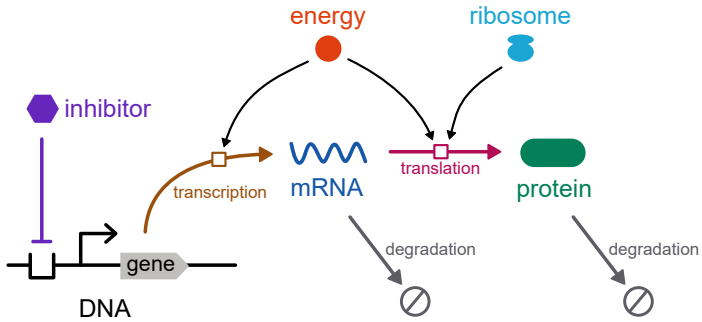

Fig. 1 Schematic of gene expression model.

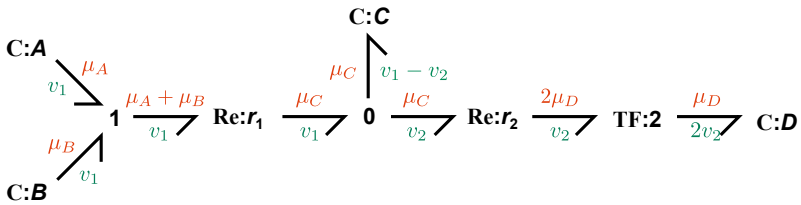

Fig. 2 The bond graph of the reaction network $A + B \rightleftharpoons C \rightleftharpoons 2D$. Each bond is labelled with a potential μ and a flow v. The components follow the notation *component type*: *name*.

We first introduce the bond graph approach for modelling biochemical systems, and explain how it ensures adherence to the laws of thermodynamics as well as its relation to kinetic modelling (Sect. 2.2). We then develop bond graph models of transcription (Sect. 2.3) and translation (Sect. 2.4). To ensure that the model of translation is computationally tractable, we apply a model reduction scheme (Sect. 2.4.2). We combine the models of transcription and translation into a model of gene expression (Sect. 2.5) and detail how models of gene expression can be composed into models of common synthetic biology motifs (Sect. 2.6).

2.2 Bond graph modelling

The bond graph is a powerful and general methodology for modelling biological systems while adhering to the laws of physics and thermodynamics. We outline the application of bond graphs to biochemistry specifically here; the reader is referred to Gawthrop and Pan (2022) for more general applications to biology.

As an illustrative example, the bond graph of the reaction network

$$A + B \rightleftharpoons C \rightleftharpoons 2D$$

is given in Fig. 2, with details explained in the following paragraphs.

Bond graphs are a generalisation of directed graphs, with physical *components* as the nodes and *bonds* as the edges. Bonds are the fundamental unit of connection in bond graphs, and are depicted using harpoons. Each bond carries two physical covariables: the chemical potential μ [J/molecule] and the molecular flow rate v

[molecules/s]. Since the product of μ and v is power [J/s], bond graphs track the movement of energy between components and ensure thermodynamic consistency.

2.2.1 Chemical species

Chemical species are labelled as **C** components because they store energy in their concentration. This can be thought of in the context of diffusion: particles in a region of higher concentration have higher potentials and therefore tend to move to regions of lower concentration. In dilute solutions, the chemical potential μ_i of a chemical species i is given by

$$\mu_i = k_B T \ln(K_i x_i),$$

where $k_B = 1.381 \times 10^{-23}$ J/K is the Boltzmann constant, T [K] is the absolute temperature, K_i [molecules^{-1}] is a thermodynamic parameter and $x_i = \int_0^t v_i(\tau) d\tau + x_i(0)$ is the molar amount of species, defined as the integral of molecular flow (Atkins and de Paula, 2006). For convenience, we define the dimensionless chemical potential as

$$\bar{\mu} = \frac{\mu_i}{k_B T} = \ln\left(K_i x_i\right). \tag{1}$$

In some cases, species are fixed at constant concentrations due to an external supply. These are represented as **Se** components whose concentrations remain constant.

2.2.2 Reactions

Reactions convert chemical species into other chemical species, and are labelled as **Re** components. Due to the principle of microscopic reversibility, all reactions are reversible in reality and therefore are modelled as reversible reactions (Polettini and Esposito, 2014). Assuming mass-action kinetics, the flux through a reaction j is given by the Marcelin-de Donder equation

$$v_j = r_j \left(e^{\mu_j^f/(k_B T)} - e^{\mu_j^r/(k_B T)} \right) = r_j \left(e^{\bar{\mu}_j^f} - e^{\bar{\mu}_j^r} \right), \tag{2}$$

where r_j is the reaction rate parameter, μ_j^f is the sum of reactant chemical potentials and μ_j^r is the sum of product chemical potentials (Gawthrop and Crampin, 2014). The bar notation indicates dimensionless chemical potentials.

2.2.3 Network topology

When a species is involved in multiple reactions, they are connected to **0** junctions. This ensures that the species accumulate at a rate equal to the sum of contributions from each reaction (mass balance). The junction also ensures that the species contributes the same chemical potential to each reaction. For example, the **0** junction in Fig. 2 ensures that the flow of the bond into the species C is equal to the difference between flows of the other two connected bonds ($\dot{x}_C = v_1 - v_2$).

When either side of a reaction involves multiple species, the species are connected to a **1** junction prior to the reaction. This junction ensures that all species are consumed or produced at the same rate, and that the contributions of the chemical potentials are summed appropriately. For example, the **1** junction in Fig. 2 ensures that the chemical potential on the outgoing bond, $\mu_A + \mu_B$, is equal to the sum of chemical potentials of incoming bonds.

2.2.4 Stoichiometry

If a species contributes multiple stoichiometry to a reaction, it is connected through a **TF** component. This component encodes the relationship

$$\mu_2 = s\mu_1,$$
$$v_1 = sv_2,$$

where s is the stoichiometry, the subscript 1 indicates the bond closer to the species and the subscript 2 indicates the bond closer to the reaction. The **TF** component ensures that the chemical potential is counted multiple times prior to calculating the contribution of a species to a reaction, and that flux is scaled up by the stoichiometric constant.

2.2.5 Relationship to kinetic modelling

A general methodology for constructing bond graphs of biochemical systems is given in Gawthrop and Crampin (2014). Using this approach, the potential balance constraints of the **1** junctions give rise to the equations

$$\mu_j^f = \sum_i s_{ij}^f \mu_i,$$
$$\mu_j^r = \sum_i s_{ij}^r \mu_i,$$

where s_{ij}^f is the stoichiometric contribution of species i to the forward side of reaction j, and similarly for s_{ij}^r on the reverse side.

By using Eqs. (1)–(2), we find that

$$v_j = r_j \left[\exp\left(\sum_i s_{ij}^f \bar{\mu}_i \right) - \exp\left(\sum_i s_{ij}^r \bar{\mu}_i \right) \right], \tag{3}$$

$$= r_j \left[\exp\left(\sum_i s_{ij}^f \ln(K_i x_i) \right) - \exp\left(\sum_i s_{ij}^r \ln(K_i x_i) \right) \right], \tag{4}$$

$$= r_j \left[\prod_i (K_i x_i)^{s_{ij}^f} - \prod_i (K_i x_i)^{s_{ij}^r} \right]. \tag{5}$$

This is the well-known law of mass action

$$v_j = k_{f,j} \prod_i x_i^{s_{ij}^f} - k_{r,j} \prod_i x_i^{s_{ij}^r}, \tag{6}$$

where the forward and reverse kinetic parameters are defined as

$$k_{f,j} = r_j \prod_{i \in \mathcal{R}_j} K_i^{s_{ij}^f},$$
$$k_{r,j} = r_j \prod_{i \in \mathcal{P}_j} K_i^{s_{ij}^r},$$

where $\mathcal{R}_j$ is the set of all reactants and $\mathcal{P}_j$ the set of all products in reaction j.

Finally, the **0** junctions ensure mass balance, thus

$$\dot{x}_i = \sum_j (s_{ij}^r - s_{ij}^f) v_j. \tag{7}$$

Eqs. (5) and (7) give rise to a set of differential equations for the biochemical system. Thus, the bond graph in Fig. 2 encodes the differential equations

$$\begin{aligned}
\dot{x}_A &= -v_1, \\
\dot{x}_B &= -v_1, \\
\dot{x}_C &= v_1 - v_2, \\
\dot{x}_D &= 2v_2,
\end{aligned}$$

where

$$\begin{aligned}
v_1 &= r_1 K_A K_B x_A x_B - r_1 K_C x_C, \\
v_2 &= r_2 K_C x_C - r_2 K_D^2 x_D^2.
\end{aligned}$$

2.3 Transcription

Following Weiße et al (2015), we model transcription as an enzyme-catalysed reaction with the energy (A) as the substrate (Sect. A.2) and mRNA (M) as the product, under the effect of an inhibitor I (such as a transcription factor). The elementary reaction network is

$$r_1 : A + E \rightleftharpoons EA \tag{8a}$$
$$r_2 : EA \rightleftharpoons E + M \tag{8b}$$
$$r_3 : E + I \rightleftharpoons EI \tag{8c}$$
$$r_4 : EA + I \rightleftharpoons EAI \tag{8d}$$
$$r_5 : EI + A \rightleftharpoons EAI \tag{8e}$$

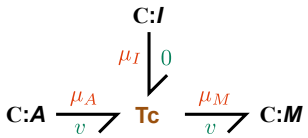

Fig. 3 Bond graph model of transcription. The **Tc** component has the rate law described in Eq. (9).

where E is the free enzyme and EA, EI and EAI are enzyme states. Under the assumption that reactions (8c)–(8d) are at quasi-equilibrium and have the same dissociation constant $K_{d,I}$, this represents non-competitive inhibition (Atkins and de Paula, 2006).

We apply the quasi-steady-state approximation (Keener and Sneyd, 2009), and therefore, the reaction network simplifies to

$$A \rightleftharpoons M$$

with the net reaction rate defined as

$$v = \bar{r}\frac{e^{\bar{\mu}_A} - e^{\bar{\mu}_M}}{(1 + \frac{e^{\bar{\mu}_A}}{R_{bA}} + \frac{e^{\bar{\mu}_M}}{R_{bM}})(1 + \frac{e^{\bar{\mu}_I}}{R_{bI}})}, \tag{9}$$

where

$$\bar{r} = \frac{r_1 r_2}{r_1 + r_2},$$

is the rate constant and

$$R_{bA} = \frac{(r_1 + r_2)K_{EA}}{r_1 K_E},$$
$$R_{bM} = \frac{(r_1 + r_2)K_{EA}}{r_2 K_E},$$
$$R_{bI} = K_I K_{d,I},$$

are the binding constants. Since the grouped parameters $\bar{r}$, R_{bA}, R_{bM} and R_{bI} are sufficient to define Eq. (9), we treat these as the parameters for transcription. Thus, Eq. (9) is represented as the three-port **Tc** component presented in Fig. 3, which is a bond graph representation of transcription that reproduces the energy-saturating behaviour in Weiße et al (2015).

While we only consider the inhibition of transcription in this study, promotors can be modelled by employing a similar approach on an enzyme activation reaction scheme (Turberville et al, 2022).

2.4 Translation

2.4.1 Full model

Because we explicitly model the use of energy and ribosomes in translation, a greater level of detail is required compared to the modelling of transcription. We first describe a detailed model, which is used as the basis for deriving a reduced model that is necessary for modelling large cell populations. Our model of translation considers the following processes:

1. **Binding** of ribosomes (R) to mRNA (M) to create a complex (C_0).
2. **Elongation**, the movement of ribosomes along the mRNA. For every site that the ribosome moves, a molecule of energy (A) is consumed, increasing the length of the nascent protein.
3. **Termination**, the release of ribosome and the completed protein (P) from the mRNA.

The reaction scheme that corresponds to these processes is:

$$M + R \underset{k_u}{\overset{k_b}{\rightleftharpoons}} C_0 \tag{10a}$$

$$C_0 + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} C_1 + M \tag{10b}$$

$$C_1 \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} C_2 \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} \dots \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} C_n \tag{10c}$$

$$C_n \underset{k_z}{\overset{k_t}{\rightleftharpoons}} R + P \tag{10d}$$

where n is the number of ATP molecules required to synthesise the protein. Consistent with the TASEP model, we assume that multiple ribosomes can bind to the same transcript, thus the mRNA molecule is released after the first elongation step (Eq. (10b)). However, for simplicity, we assume that apart from the first binding site, ribosomes do not obstruct each other, so that the complexes may be treated independently. While both binding and termination kinetics are dependent on the mRNA sequence (Salis, 2011; Prabhakar et al, 2023), we assume for simplicity that both initiation and termination dynamics are independent of sequence. A bond graph of the full model is shown in Fig. 4a.

2.4.2 Model reduction

Since the number of ATP molecules per protein n can be on the order of thousands, it is computationally expensive to use the full model of translation in models of complex gene regulatory networks. Here, we outline a method for simplifying the full model while retaining its essential features. We group the complexes C_1 to C_{n-1} into a

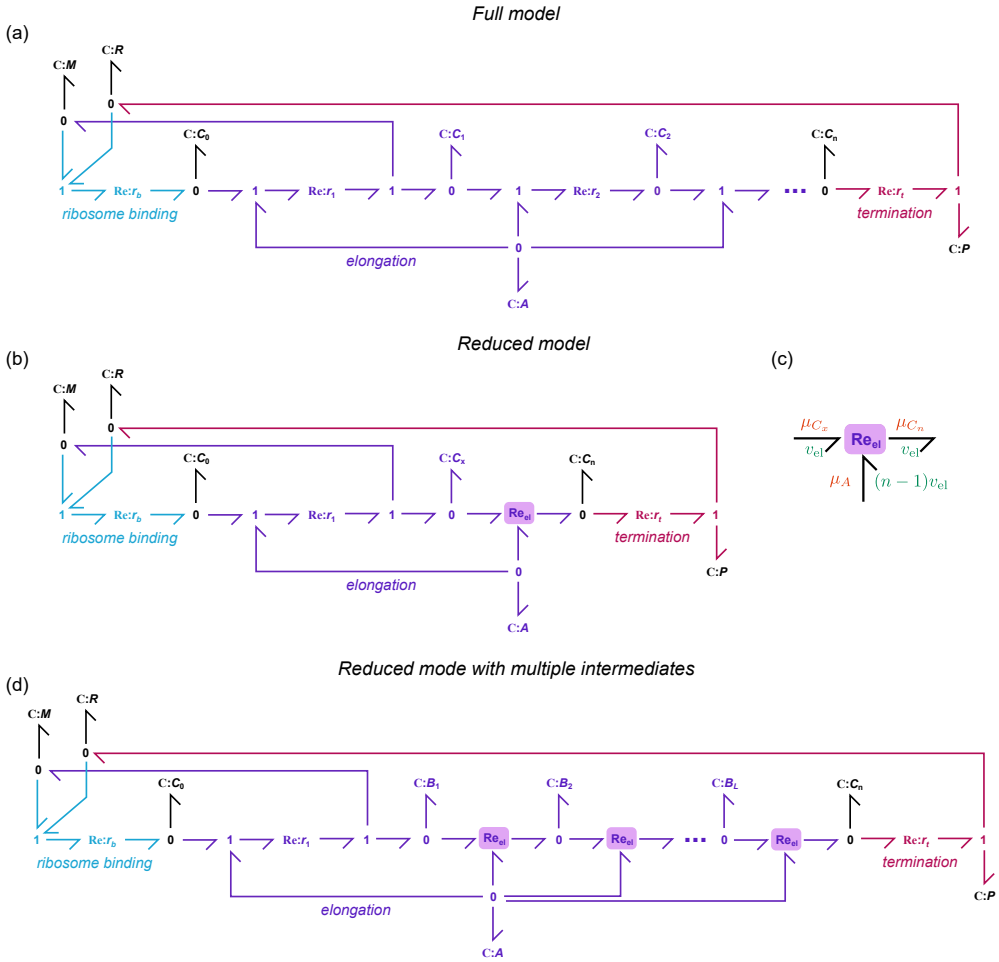

Fig. 4 Bond graph models of translation. (a) Bond graph of the full translation model (Sect. 2.4.1); (b) bond graph of the reduced model (Sect. 2.4.2); (c) the $\mathrm{Re}_{\mathrm{el}}$ component used to model elongation in (b); (d) bond graph of a reduced model with multiple intermediates.

lumped complex C_x, giving rise to the reaction scheme

$$M + R \underset{k_u}{\overset{k_b}{\rightleftharpoons}} C_0 \underset{k_r M}{\overset{k_f A}{\rightleftharpoons}} C_x \overset{v_{\mathrm{el}}}{\rightleftharpoons} C_n \underset{k_z}{\overset{k_t}{\rightleftharpoons}} R + P, \tag{11}$$

where v_{el} is a non-mass-action reaction representing the elongation rate. A bond graph of the system is shown in Fig. 4b, where the elongation reaction is represented as a three-port reaction component (Fig. 4c). We aim to find an expression for v_{el} under a quasi-steady-state approximation. Under this choice of v_{el}, the steady-state elongation rate is consistent with the full model.

To simplify elongation, we consider the reactions associated with elongation after the first step:

$$C_1 \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} C_2 \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} \cdots \underset{k_r}{\overset{k_f A}{\rightleftharpoons}} C_n. \tag{12}$$

To establish a connection to the reactions involved in translation, we make the simplifying assumption that C_1 and C_n are boundary species whose concentrations change slowly relative to the other complexes in the chain. We expect this assumption to be reasonable since binding dynamics are generally slower than elongation dynamics. Nonetheless, we outline a method in Sect. 2.4.5 for achieving a better match in cases where this assumption does not hold. Following the quasi-steady-state approximation, $\mathrm{d}C_i/\mathrm{d}t = 0$ for $i \in \{2, \ldots, n-1\}$. Then

$$\frac{\mathrm{d}C_i}{\mathrm{d}t} = k_f A C_{i-1} + k_r C_{i+1} - (k_f A + k_r) C_i = 0, \quad i \in 2, \ldots, n-1. \tag{13}$$

To solve this system of equations, we treat C_i as a discrete sequence and analyse its z-transform. For simplicity, we define $x_i = C_{i+1}$ for $i = 0, \ldots, n-1$ and $X(z)$ as the z-transform of x_i. The sequence follows the difference equation

$$k_r x_{i+2} - (k_f A + k_r) x_{i+1} + k_f A x_i = 0, \quad i \geq 0.$$

Taking the z-transform of both sides, we find

$$k_r z^2 X(z) - (k_f A + k_r) z X(z) + k_f A X(z) = p(z),$$

where $p(z)$ is a quadratic used to initialise the sequence. We then factorise the quadratic on the left-hand side:

$$k_r (z-1)(z - \hat{A}) X(z) = p(z),$$

where $\hat{A} = k_f A / k_r$. Using partial fraction decomposition,

$$X(z) = \frac{az}{z-1} + \frac{bz}{z - \hat{A}},$$

for some constants a and b. Hence, by taking the inverse z-transform,

$$C_i = x_{i-1} = a + b\hat{A}^{i-1}, \quad i \geq 1. \tag{14}$$

The constants a and b can be solved for substituting the boundary values C_1 and C_n:

$$a = -\frac{C_1 \hat{A}^{n-1} - C_n}{1 - \hat{A}^{n-1}},$$

$$b = \frac{C_1 - C_n}{1 - \hat{A}^{n-1}}.$$

Hence, the elongation rate is

$$v_{\text{el}} = k_f A C_i - k_r C_{i+1} = k_r a(\hat{A} - 1) = k_r \left(C_1 \hat{A}^{n-1} - C_n \right) \frac{1 - \hat{A}}{1 - \hat{A}^{n-1}}. \tag{15}$$

We make a few remarks on this expression:

1. In the case where the lumped complex contains a single species ($n = 2$), the elongation rate reduces to the law of mass action.
2. If $\hat{A} \ll 1$, the elongation rate is well approximated by the law of mass action:

$$v_{\text{el}} \approx k_r \left(C_1 \hat{A}^{n-1} - C_n \right).$$

3. If $\hat{A} \gg 1$ and $C_n \ll \hat{A}^n C_1$, the elongation rate is an approximately linear function of $\hat{A}$:

$$v_{\text{el}} \approx k_r C_1 (\hat{A} - 1).$$

4. When n is high and $C_n \ll \hat{A}^n C_1$, the elongation rate is approximately a piecewise linear function:

$$v_{\text{el}} \approx \begin{cases} k_r C_n (\hat{A} - 1), & \hat{A} < 1, \\ k_r C_1 (\hat{A} - 1), & \hat{A} \geq 1. \end{cases} \tag{16}$$

Interestingly, this behaviour is analogous to rectified conductances seen in ion channels (Keener and Sneyd, 2009).

2.4.3 Bond graph component

To use the reduced model in a bond graph, the kinetic parameters in Eq. (15) need to be expressed in terms of bond graph parameters. We note that since the kinetic parameters k_f and k_r are identical between elongation steps, $K_{C_{i+1}}/K_{C_i} = \alpha$ for some constant α. We also note that from Eq. (1), $K_i x_i = e^{\bar{\mu}_i}$. Then

$$k_r C_1 = r_n K_{C_n} x_{C_1} = r_n \frac{K_{C_n}}{K_{C_1}} K_{C_1} x_{C_1} = r_n \alpha^{n-1} e^{\bar{\mu}_{C_1}}, \tag{17}$$

$$k_r C_n = r_n K_{C_n} x_{C_n} = r_n e^{\bar{\mu}_{C_n}}, \tag{18}$$

$$\hat{A} = \frac{k_f A}{k_r} = \frac{r_i K_{C_i} K_A x_A}{r_i K_{C_{i+1}}} = e^{\bar{\mu}_A} \alpha^{-1}, \tag{19}$$

where r_i is the rate parameter for the reaction $C_{i-1} + A \rightleftharpoons C_i$.

Therefore, defining $m = n - 1$, we can express Eq. (15) as

$$v_{\text{el}} = \left(r_n e^{\bar{\mu}_{C_1}} e^{m\bar{\mu}_A} - r_n e^{\bar{\mu}_{C_n}}\right) \frac{1 - e^{\bar{\mu}_A}\alpha^{-1}}{1 - e^{m\bar{\mu}_A}\alpha^{-m}} \tag{20}$$

$$= r\left(e^{\bar{\mu}_{C_1}} - e^{\bar{\mu}_{C_n} - m\bar{\mu}_A}\right) \frac{1 - e^{\bar{\mu}_A}\alpha^{-1}}{(\alpha e^{-\bar{\mu}_A})^m - 1}, \tag{21}$$

where $r = r_n\alpha^m$. Note that Eq. (21) is a thermodynamically consistent expression as $v > 0$ only if $\mu_{C_1} + m\mu_A - \mu_{C_n} > 0$. Therefore, the reversible model of elongation can be encoded as a three-port bond graph component (Fig. 7c) with the constitutive relations

$$\begin{aligned} v &= r\left(e^{\bar{\mu}_{C_1}} - e^{\bar{\mu}_{C_n} - m\bar{\mu}_A}\right) \frac{1 - e^{\bar{\mu}_A}\alpha^{-1}}{(\alpha e^{-\bar{\mu}_A})^m - 1}, \\ v_{C_1} &= -v, \\ v_{C_n} &= v, \\ v_A &= -mv. \end{aligned}$$

Thus, while the bond graph model of translation has been substantially reduced in complexity, it retains the thermodynamic consistency of the full model.

2.4.4 Lumped complex

The lumping scheme in Sect. 2.4.2 groups several ribosomal complexes into the same species. Accordingly, the initial condition of the lumped complex C_x is set to

$$C_x(t = 0) = \sum_{i=1}^{n-1} C_i(t = 0).$$

Assuming that $k_f A \gg k_r$, all complexes C_i have similar concentrations at steady state, and hence

$$C_x \approx (n - 1)C_i.$$

Due to the difference in amount between C_x and C_i, the thermodynamic constant of C_x is adjusted to maintain the same reaction rates between full and reduced models:

$$K_{C_x} = \frac{K_{C_1}}{n - 1}.$$

2.4.5 Multiple lumped intermediates

As a generalisation to the earlier lumping scheme, one may lump complexes into more than one intermediate if greater accuracy is required. With L lumps of length k, the

reaction scheme simplifies to

$$M + R \underset{k_u}{\overset{k_b}{\rightleftharpoons}} C_0 \underset{k_r M}{\overset{k_f A}{\rightleftharpoons}} B_1 \overset{v_{\mathrm{el},1}}{\rightleftharpoons} B_2 \overset{v_{\mathrm{el},2}}{\rightleftharpoons} \dots B_L \overset{v_{\mathrm{el},l}}{\rightleftharpoons} C_n \underset{k_z}{\overset{k_t}{\rightleftharpoons}} R + P, \tag{22}$$

where B_i are groups of k_i complexes and $v_{\mathrm{el},i}$ is the net rate of the i-th lumped elongation reaction, following the rate law Eq. (21) with $m = k_i$. The number of complexes in lump B_i is

$$k_i = q + \mathbf{I}(i < r + 1),$$

where $q = \lfloor (n-1)/L \rfloor$ is the integer quotient when $n-1$ is divided by L with remainder $r = (n-1) - qL$, and $\mathbf{I}$ is the indicator function. This ensures consistency in the number of complexes (i.e. $\sum_{i=1}^{L} k_i = n - 1$) while keeping the size of each lump as consistent as possible. A bond graph of this reduced model is shown in Fig. 4d.

2.5 Gene expression and regulation

We construct a model of gene expression and regulation by coupling together the earlier models of transcription and translation (Fig. 5). Since bond graphs are graphical by nature, they enable a modular approach to modelling biological systems. Using this approach, models of subsystems – in this case the transcription and translation modules – can be re-used in larger models without the need to manually re-derive the equations. Here, we couple bond graph modules using the approach in Gawthrop and Pan (2022), where **C** components corresponding to shared species are exposed as ports within the transcription and translation modules.

In addition to transcription and translation, degradation processes (Fig. 5c) are added for mRNA and proteins to avoid overaccumulation. In biology, both mRNA and proteins are known to be enzymatically degraded for regulatory purposes (Garneau et al, 2007; López-Otín and Overall, 2002). Degradation rates can vary between mRNA and proteins, depending on their sequence, and are a major regulator of protein abundance and the timescale of their dynamics (Elowitz and Leibler, 2000). These involve the chemical transformation of such species into a sink X, representing the degradation products (e.g. nucleotides and amino acids). The degradation rate of a substrate S is

$$v_{\mathrm{deg},S} = r_{\mathrm{deg},S}\left(e^{\bar{\mu}_S} - e^{\bar{\mu}_X}\right) = r_{\mathrm{deg},S} K_S x_S - r_{\mathrm{deg},S} e^{\bar{\mu}_X}.$$

While it is physically impossible for chemical potentials to be infinite, the chemical potential $\bar{\mu}_X$ is set to a very negative value so that

$$v_{\mathrm{deg}} \approx r_{\mathrm{deg},S} K_S x_S,$$

and hence the decay rate of the molecule is $r_{\mathrm{deg},S} K_S$ [s^{-1}].

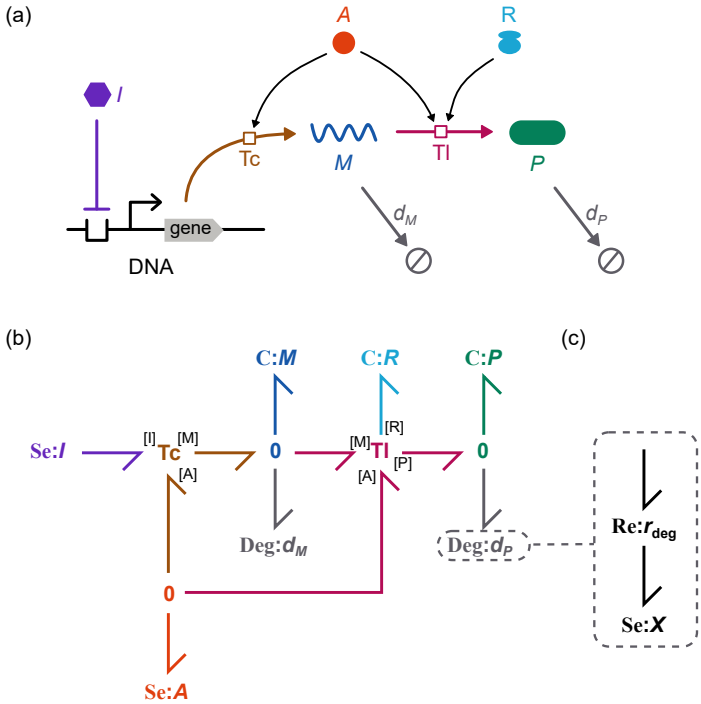

Fig. 5 Model of gene expression. (a) Schematic and (b) bond graph model of gene expression. Degradation for mRNA and protein are modelled using the modules in (c). The transcription (Tc) module is shown in Fig. 3 and the translation (Tl) module is shown in Fig. 4b.

The differential equations implied by the bond graph in Fig. 5b are

$$\begin{aligned}
\frac{\mathrm{d}x_M}{\mathrm{d}t} &= v_{\mathrm{Tc}}(x_A, x_M, \mu_I) - v_b(x_M, x_R, x_{C_0}) - v_{\mathrm{deg},M}(x_M) + v_1(x_{C_0}, x_A, x_{C_x}, x_M),\\
\frac{\mathrm{d}x_R}{\mathrm{d}t} &= -v_b(x_M, x_R, x_{C_0}) + v_t(x_{C_n}, x_P, x_R),\\
\frac{\mathrm{d}x_P}{\mathrm{d}t} &= v_t(x_{C_n}, x_P, x_R) - v_{\mathrm{deg},P}(x_P),\\
\frac{\mathrm{d}x_{C_0}}{\mathrm{d}t} &= v_b(x_M, x_R, x_{C_0}) - v_1(x_{C_0}, x_A, x_{C_x}, x_M),\\
\frac{\mathrm{d}x_{C_x}}{\mathrm{d}t} &= v_1(x_{C_0}, x_A, x_{C_x}, x_M) - v_{\mathrm{el}}(x_{C_x}, x_{C_n}, x_A),\\
\frac{\mathrm{d}x_{C_n}}{\mathrm{d}t} &= v_{\mathrm{el}}(x_{C_x}, x_{C_n}, x_A) - v_t(x_{C_n}, x_P, x_R),
\end{aligned}$$

where the reaction fluxes are

$$v_{\mathrm{Tc}} = \bar{r}_{\mathrm{Tc}} \frac{K_A x_A - K_M x_M}{(1 + \frac{K_A x_A}{R_{bA}} + \frac{K_M x_M}{R_{bM}})(1 + \frac{e^{\bar{\mu}_I}}{R_{bI}})}, \tag{23}$$

$$v_b = r_b(K_M K_R x_M x_R - K_{C_0} x_{C_0}), \tag{24}$$

$$v_1 = r_1(K_{C_0}K_A x_{C_0}x_A - K_{C_x}K_M x_{C_x}x_M), \tag{25}$$

$$v_{\text{el}} = r_{\text{el}}\left(K_{C_x}K_A^{(n-1)}x_{C_x}x_A^{(n-1)} - K_{C_n}x_{C_n}\right)\frac{1 - K_A x_A \alpha^{-1}}{1 - \left(\frac{K_A x_A}{\alpha}\right)^{n-1}}, \tag{26}$$

$$v_t = r_t(K_{C_n}x_{C_n} - K_P K_R x_P x_R), \tag{27}$$

$$v_{\text{deg},M} = r_{\text{deg},M}K_M x_M - r_{\text{deg},M}e^{\bar{\mu}_{X,M}}, \tag{28}$$

$$v_{\text{deg},P} = r_{\text{deg},P}K_P x_P - r_{\text{deg},P}e^{\bar{\mu}_{X,P}}. \tag{29}$$

2.6 Synthetic biology motifs

Certain common synthetic biology motifs can be modelled by coupling together multiple instances of the gene expression model in Sect. 2.5. To study the effects of resource availability on synthetic circuits, we construct models for two well-known motifs in synthetic biology: (a) the toggle switch, comprising two proteins that mutually inhibit each other (Fig. 6a–c); and (b) the repressilator, consisting of three proteins that inhibit each other in a cycle (Fig. 6d–f). Bond graphs of these modules are shown in Fig. 6. These consist of an instance of the gene expression module **GE** for each gene, but with additional connections between the genes. In particular, the protein of one gene expression system acts as the repressor for another gene expression system. These connections are shown by the coloured bonds in Fig. 6. Also, since all genes require energy and ribosomes to be expressed, these resources must be shared between modules. This is modelled by connecting the modules to **0** junctions (grey bonds in Fig. 6) to ensure mass conservation.

Gene regulatory networks often require binding cooperativity to enhance their sensitivity to inputs (Keener and Sneyd, 2009). This is incorporated by using a **TF** component (Sect. 2.2.4) with modulus h to model the binding of multiple inhibitors to prevent transcription, resulting in a transcription rate of

$$v_{\text{Tc}} = \bar{r}_{\text{Tc}}\frac{K_A x_A - K_M x_M}{(1 + \frac{K_A x_A}{R_{bA}} + \frac{K_M x_M}{R_{bM}})(1 + \frac{K_I^h x_I^h}{R_{bI}})},$$

A Hill coefficient of $h = 2$ is used, consistent with previous models (Elowitz and Leibler, 2000; Weiße et al, 2015).

3 Results

3.1 Comparison of full and reduced models

3.1.1 Elongation

To ensure the validity of the model reduction scheme in Sect. 2.4.2, we use dynamic simulations to compare the output of the full and reduced models of translation. We first examine the elongation reactions in isolation to verify that the simple model matches the elongation rate of the full model at steady state. Fig. 7a shows the results obtained from simulating the full model of elongation (Eq. (12)), compared to the analytical form of the elongation rate (Eq. (15)). For simplicity, we use a chain length

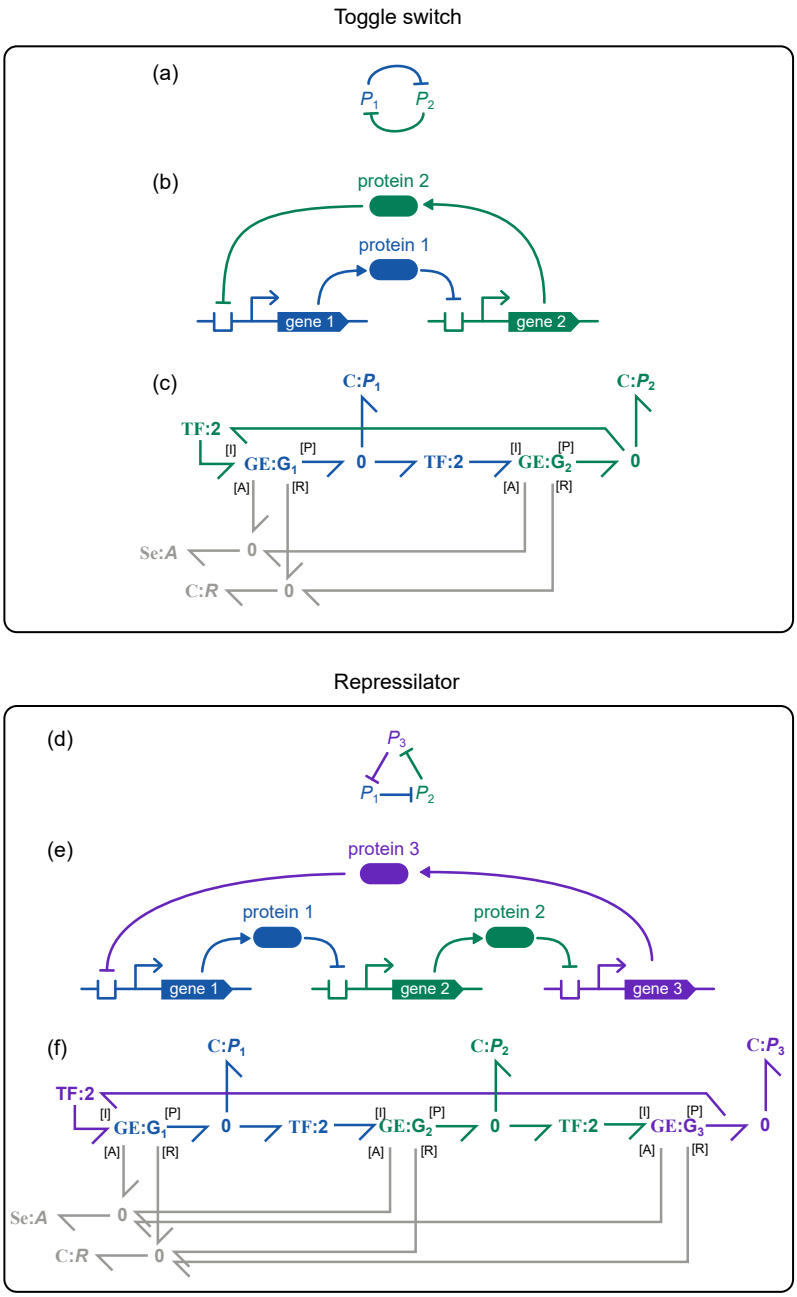

Fig. 6 Models of synthetic biology motifs. The bond graph model of (a–c) the toggle switch; and (d–f) repressilator. For each panel, the relationships between proteins are shown on top (a,d), the mechanisms of interaction are displayed in the middle (b,e) and bond graphs are shown on the bottom (c,f). The **GE** modules refer to the model in Fig. 5b.

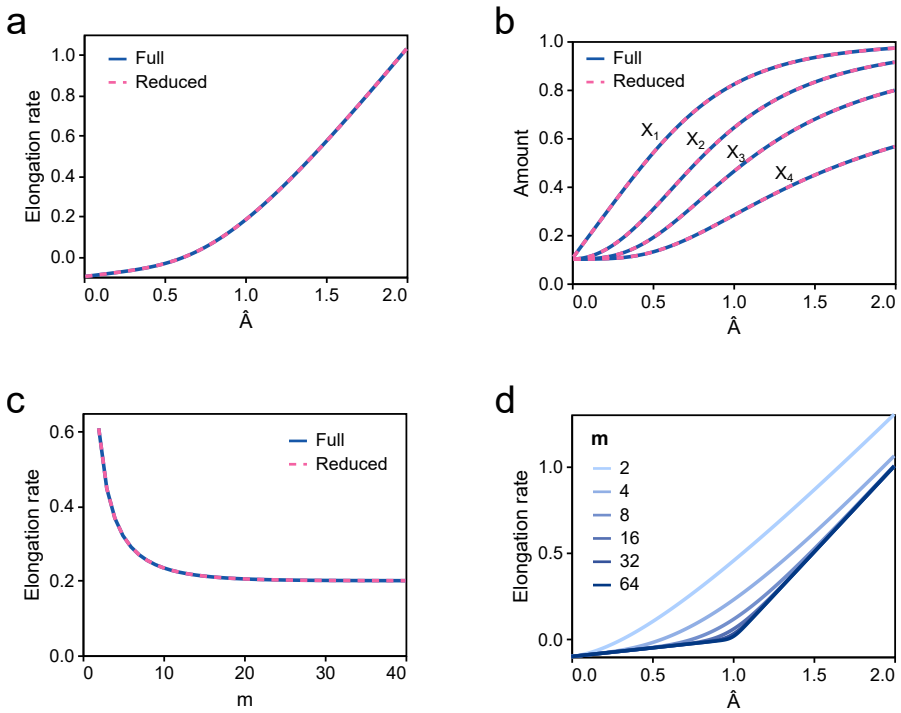

Fig. 7 A simplified model of elongation. Steady-state comparisons for (a) elongation rates and (b) complex amounts are plotted for different values of $\hat{A}$. (c) Steady-state comparison of elongation rates for different values of m. The simulations in (a)–(c) use the kinetic parameters $k_f = 1$, $k_r = 1$, $C_0 = 1.0$ and $C_n = 0.1$, with A set to 1.2 for (c). (d) A plot of the analytical solution for different values of m. Models are run for $t = 50$ minutes to achieve steady state.

of $m = 5$ for these simulations, but the results generalise to longer chain lengths (see below). As expected, the full and reduced models match very well. The steady-state amounts of the intermediate complexes X_1, X_2, X_3 and X_4 are plotted in Fig. 7b, and also match closely. We next investigate whether the model reduction scheme holds for different chain lengths m (Fig. 7c). The elongation rate decreases, and ultimately plateaus at high m; and the reduced model is able to predict this behaviour. Thus, the reduced model is able to accurately match the steady-state behaviour of the full model under a wide range of conditions.

Fig. 7d shows the relationship between elongation rate (Eq. (15)) and $\hat{A}$ for different chain lengths m. For shorter chain lengths, the elongation rate is smooth with respect to $\hat{A}$. However, as predicted by equation Eq. (16), for large m, the curve is almost piecewise linear with different slopes on either side of $\hat{A} = 1$.

3.1.2 Translation

While the above results confirm similarities between the full and reduced models at steady state, we are interested in comparing their behaviours when included in a dynamic model. Thus, we next investigate whether the dynamics of translation are

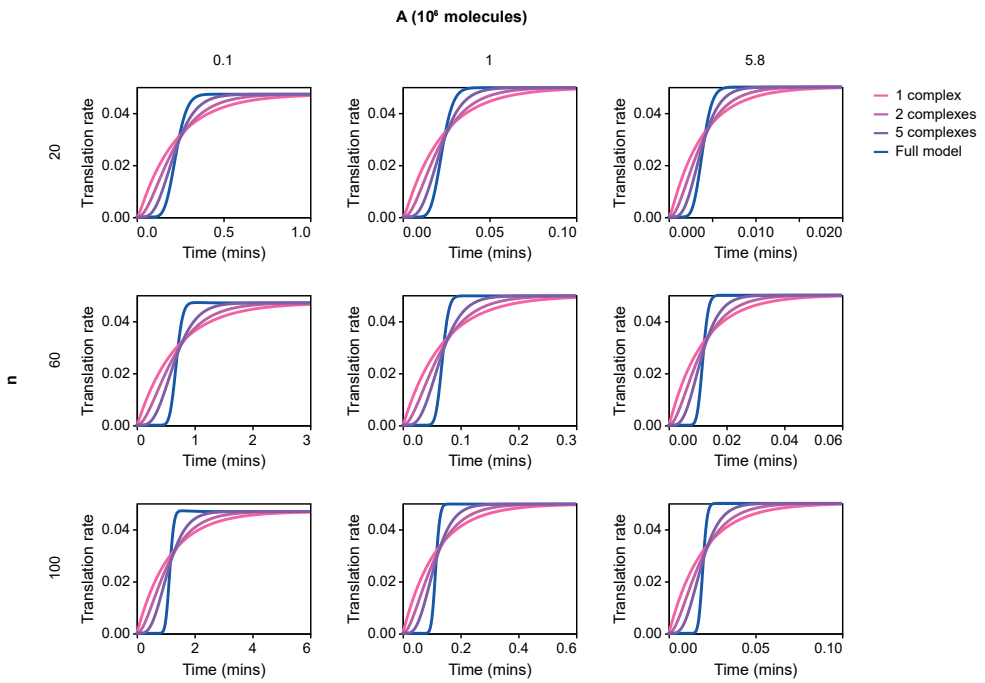

Fig. 8 Comparison of the dynamics of translation rate. The energetic state A is varied horizontally, and the chain length n is varied vertically. The translation rate is given in proteins per minute. Simulations are run with 10 mRNA molecules, 5000 ribosomes and $\bar{\mu}_P = -15$ to approximate the immediate removal of proteins.

similar between the full and reduced models. To simulate the models under biological conditions, the models are parameterised based on previous experimental and modelling work (see Appendix A). We simulate the dynamics of the full (Fig. 4a) and reduced models (Fig. 4b), and compare them for different energy levels and protein lengths (Fig. 8). Models are initialised from an inactivated state with very small amounts of ribosome-mRNA complexes, and the evolution to an activated state is compared between models. We test the match between models for a range of energetic states A, with $A = 5.8 \times 10^6$ molecules/cell corresponding most closely to an energy-rich environment (Sect. A.1). We also test the quality of match for different chain lengths n.

In all cases, the reduced model approaches the same translation rate as the full model. However, the models exhibit different dynamics. Whereas the full model has a more step-like response to a sudden increase in mRNA, the reduced model approaches the steady-state translation rate more gradually. As one would expect, the match between dynamics becomes poorer for larger values of n. Nonetheless, both the full and reduced models approach their steady states at similar timescales for a wide range of chain lengths. Due to the similarities in steady-state behaviour and timescales, we believe that the simple model is an adequate approximation, and we indeed show

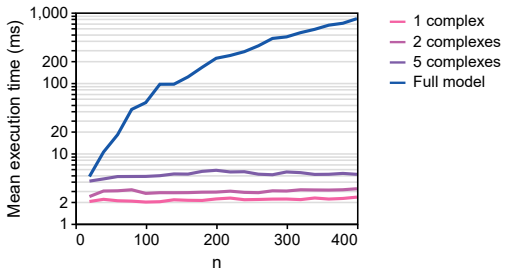

Fig. 9 Benchmark of full and reduced models. Simulations are run on the translation model and the mean execution time is plotted for different protein lengths n. Models were solved using the CVODE Backward Differentiation Formula method (Hindmarsh et al, 2005) with a relative tolerance of 10^{-9} and absolute tolerance of 10^{-12}. Models were repeatedly simulated over either 1000 samples, or until 1 minute had elapsed in real time. The code was simulated on a Windows 10 server with 4 CPU cores (AMD EPYC 7702 processor).

in Sect. 3.3 that the differences in dynamics result in very minor differences at the level of gene expression. In circumstances where better matching with the dynamic behaviour is required, one could lump elongation into multiple intermediate states rather than a single complex. This approach is analogous to epidemiological models with multiple exposed or infectious compartments (Hurtado and Kirosingh, 2019) and models of multi-stage transport processes (Dowling et al, 2022). Simulations of two and five lumped complexes are displayed in Fig. 8, and show a gradual improvement of fit with the number of intermediates.

We also find that the reduced model is computationally efficient compared to the full model. A benchmark of runtimes is presented in Fig. 9. As expected, the full model takes longer to run as the number of state variables increases; the approximately linear gradient at high n indicates a rapidly increasing runtime with protein length. In contrast, the time taken to run the reduced model remains stable even for higher protein lengths. At the longest tested protein length of $n = 400$, a 300× speedup is observed for the reduced model, relative to the full model. We expect even greater speedup for realistic protein lengths of over 1000 ATP molecules per protein.

3.2 Energetic analysis

A strength of the bond graph approach is that it can reveal insights into the thermodynamic efficiency of biological processes. We examine the energetics of translation here. Translation can be seen as the use of energy from ATP hydrolysis to form peptide bonds in proteins, with the overall reaction

$$nA \rightleftharpoons P.$$

Thus, we can define the thermodynamic efficiency η of translation as the proportion of total energy from ATP hydrolysis used to form proteins:

$$\eta = \frac{\bar{\mu}_P}{\bar{\mu}_A}.$$

Table 1 Energetic analysis of translation. Models of translation with $n = 1200$ are simulated with different amounts of energy. The values recorded are the translation rate (v_{trans} [molecules/s]), energy supplied ($\bar{\mu}_A$), chemical potential of protein ($\bar{\mu}_P$), thermodynamic efficiency (η), free energy of ribosomal binding ($\Delta G_{\text{binding}} = \bar{\mu}_{C_0} - \bar{\mu}_M - \bar{\mu}_R$), free energy of elongation ($\Delta G_{\text{elongation}} = \bar{\mu}_{C_n} + \bar{\mu}_M - \bar{\mu}_{C_0} - n\bar{\mu}_A$) and free energy of termination ($\Delta G_{\text{termination}} = \bar{\mu}_R + \bar{\mu}_P - \bar{\mu}_{C_n}$).

A	v_{trans}	μ_A	μ_P	η (%)	$\Delta G_{\text{binding}}$	$\Delta G_{\text{elongation}}$	$\Delta G_{\text{termination}}$
1×10^5	42.1	15.9	1486.9	7.77	−11.4	−17629.2	-9.16×10^{-4}
1×10^6	49.0	18.2	1486.9	6.79	−13.7	−20390.0	-9.57×10^{-4}
5.8×10^6	49.8	20.0	1486.9	6.20	−15.4	−22497.7	-9.61×10^{-4}

We simulate the translation model for different values of A and calculate the thermodynamic efficiency for each of these energetic states (Table 1). The thermodynamic efficiency of translation is 6–8%, which is low compared to many other biological processes (Pan et al, 2019). While this may be surprising at first glance, the apparently wasteful nature of translation has previously been attributed to kinetic proofreading, a mechanism for reducing the error rate (Hopfield, 1974). While increasing A increases the translation rate v_t, it also reduces the efficiency since the input potential μ_A increases without a corresponding increase in the output potential μ_P.

The thermodynamic analysis in Table 1 also allows the thermodynamic consistency of the model to be verified. The free energies of binding ($M + R \rightleftharpoons C_0$), elongation ($C_0 + nA \rightleftharpoons C_n + M$) and termination ($C_n \rightleftharpoons R + P$) are defined as $\Delta G_{\text{binding}}$, $\Delta G_{\text{elongation}}$ and $\Delta G_{\text{termination}}$ respectively. One can easily verify that energy balance is satisfied, i.e. $\bar{\mu}_P - n\bar{\mu}_A = \Delta G_{\text{binding}} + \Delta G_{\text{elongation}} + \Delta G_{\text{termination}}$. The second law of thermodynamics is also satisfied since all the free energies are negative.

The free energies also provide a more detailed picture of how power is dissipated. The vast majority of energy is dissipated during elongation. While the binding free energy is much lower by comparison, it is still far from equilibrium. On the other hand, due to the fast nature of the termination reaction, it effectively operates at equilibrium ($\Delta G_{\text{elongation}} \approx 0$). Thus, it is likely that both the binding and elongation rates control the rate of protein translation, which is consistent with current hypotheses (Raveh et al, 2016; Jacques and Dreyfus, 1990).

3.3 Gene expression

Models of gene expression are constructed by coupling models of transcription and translation (Fig. 5). We simulate the induction of a gene, where an inhibitor is initially present at high concentrations before being removed (Fig. 10a). As expected, following the removal of the inhibitor, mRNA levels rise followed by a rise in the the protein. The counts of these molecules are within the ranges observed experimentally (Milo, 2015). In Sect. A.5.1, we show that under the assumption of excess energy, the steady-state mRNA and protein numbers are

$$x_M = \frac{1}{r_{\text{deg},M} K_M} \frac{\bar{r}_{\text{Tc}} R_{bA}}{1 + e^{\bar{\mu}_I}/R_{bI}},$$

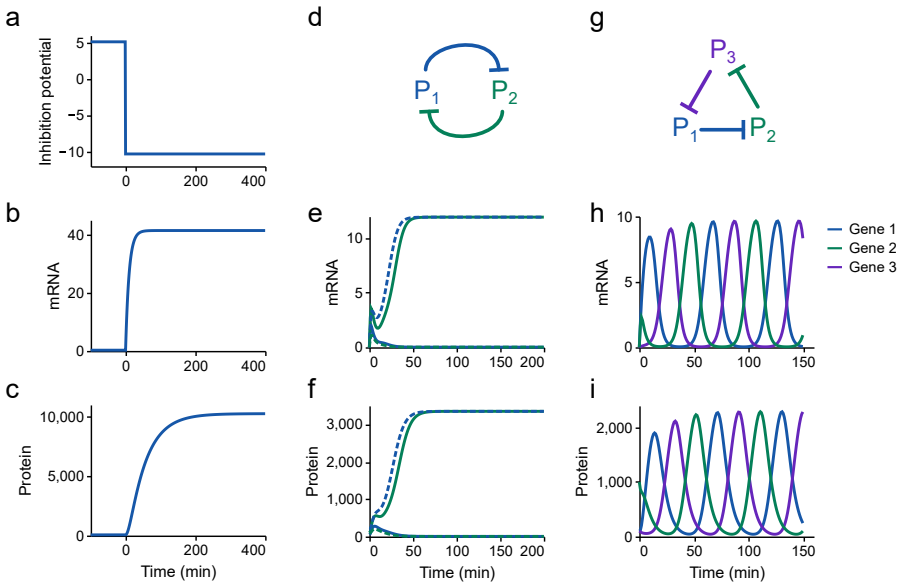

Fig. 10 Simulations of synthetic circuits. (a–c) Simulations of a single gene. The inhibition potential $\mu_I - \ln(R_{bI})$ is reduced from a high to low value at $t = 0$, indicating a switch from an inactive to active state. (d–f) Simulations of the toggle switch. Two simulations (represented using solid and dashed lines) are run with different initial conditions. The steady states are dependent on the initial condition, indicating bistability. (g–i) Simulations of the repressilator circuit.

$$x_P = \frac{r_b K_M K_R x_R}{r_{\mathrm{deg},M} K_M r_{\mathrm{deg},P} K_P} \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + e^{\bar{\mu}_I}/R_{bI}}.$$

With the parameters in the model, $x_M \approx 41$ and $x_P \approx 10351$, which is in good agreement with Fig. 10b–c. The power dissipation rate is plotted against time in Fig. 18, which indicates that protein synthesis and degradation account for the majority of energy dissipated within the circuit.

To ensure that the model reduction scheme in Sect. 2.4.2 is valid, the same models are simulated with the full model of translation (Fig. 4a) in place of the reduced models. The resulting mRNA and protein levels are virtually identical to that of Fig. 10 (Fig. 11), and the quality of match is similar for the models of the toggle switch and repressilator. Thus, we expect that translation operates near steady state under biological conditions, and that the loss of dynamics from model reduction is unimportant under these conditions.

In Fig. 12, the translation rate is plotted against the amount of energy and ribosomes. Decreasing the amount of energy has little effect on the translation rate (Fig. 12a) until the amount of energy is very low, reflecting energy starvation conditions. The translation rate reaches 50% of its maximum value at $A = 5.63 \times 10^4$ molecules (Fig. 12a; inset), which corresponds to a concentration of $K_m^{\mathrm{ATP}} = 93\ \mu$M. It is reassuring that this value is on the same order of magnitude as experimental measurements ($K_m^{\mathrm{ATP}} \approx 27\ \mu$M; Jewett et al (2009)) despite the fact that the model has not been fitted to such data. Decreasing the number of ribosomes reduces the

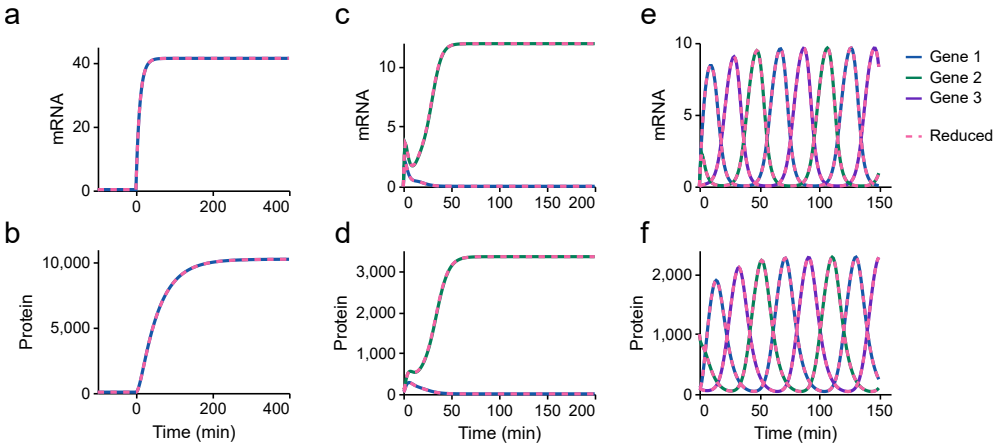

Fig. 11 Comparison of full and reduced models of synthetic circuits. Models of synthetic circuits using the full translation model (solid lines) are compared to the corresponding models using the reduced translation model (dashed lines). (a–b) Induction of a single gene; (c–d) toggle switch; (e–f) repressilator.

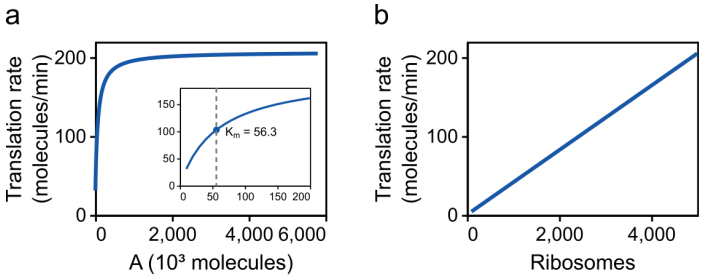

Fig. 12 The effect of resource abundance on translation rate. The translation rate is plotted against (a) amount of energy; and (b) the number of ribosomes. The inset in (a) shows the translation rate for low amounts of energy.

translation rate linearly (Fig. 12b). This appears to be consistent with studies that hypothesise ribosomes are a limiting factor for translation (Vind et al, 1993). We also show in Sect. A.5.1 that the steady-state translation rate is linearly dependent on ribosome number at high energy levels.

3.4 Resource dependence of synthetic circuits

3.4.1 Toggle switch

We next explore the behaviours of synthetic biology circuits when modelled with thermodynamic dependencies in mind. A model of the toggle switch (Fig. 6a–c) is constructed by coupling together two mutually inhibiting genes. We perform simulations of the toggle switch under two different initial conditions (Fig. 10d–f), corresponding to the activation of protein 1 (dashed lines) and protein 2 (solid lines). As expected, the

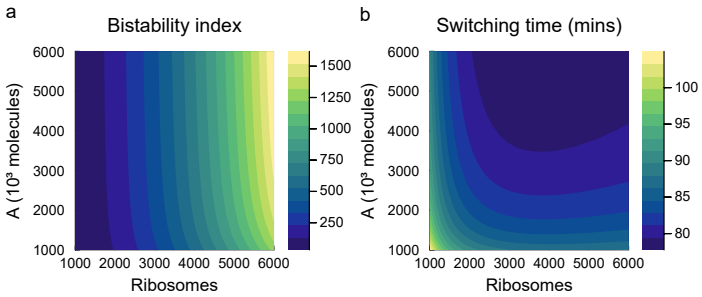

Fig. 13 Simulations of the toggle switch under varying energy and ribosome levels. Simulations are run with an initially high amount of protein for gene 1, resulting in a steady state favouring this gene. (a) The bistability index, defined as the steady-state amount of protein 1 in its active state relative to its inactive state; (b) switching time, defined as the time required for the system to be within 1% of its steady-state value (Eq. (30)).

model reproduces the multistability of the switch observed experimentally (Gardner et al, 2000).

As discussed in Sect. 1, synthetic circuits do not operate in isolation, and are affected by background processes such as the allocation of energy and ribosomes. We examine the effects of limiting either energy (A) or ribosome (R) availability on the performance of the toggle switch. The performance of the toggle switch is characterised using two measures:

1. the *bistability index*, defined as the steady-state amount of P_1 in the active steady-state relative to its amount in the inactive state (Chen, 2006).
2. the *switching time*, defined as the time taken for both proteins to reach their steady state (Gardner et al, 2000). To quantify the switching time, we define the distance from the steady state as

$$d = \sqrt{\left(\frac{P_1 - P_1^*}{P_1^*}\right)^2 + \left(\frac{P_2 - P_2^*}{P_2^*}\right)^2}, \tag{30}$$

where P_1^* and P_2^* are the steady-state amounts of proteins 1 and 2 respectively. The switching time is the time taken for the model to reach a distance of $d < 0.01$.

Fig. 13 shows the effect of energy and ribosomes on the behaviour of the toggle switch. We find that ribosomes are the primary determinant of toggle switch bistability, as the bistability index varies by approximately 30-fold over the range of ribosome amounts considered. The steady states of the toggle switch under energy-rich conditions are analysed in Sect. A.5.2, verifying this result. In contrast, both the energy and ribosomes affect the switching time. A higher energy availability reduces the switching time as expected, since energy accelerates the dynamics of translation. Surprisingly, the number of ribosomes has a nonmonotonic effect on switching time, and achieves a minimum value between 3000–6000 ribosomes. This leads to potential trade-offs in performance at high numbers of ribosomes; increasing ribosomal availability increases the bistability index, but also increases the switching time.

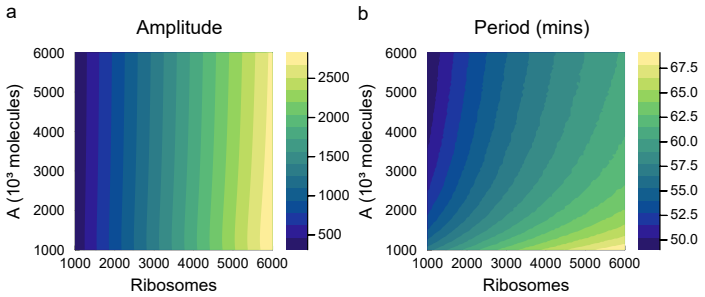

Fig. 14 Simulations of the repressilator under varying energy and ribosome levels. Simulations are run until a limit cycle is reached, and the (a) amplitude and (b) period of oscillations are recorded.

3.4.2 Repressilator

We next study the dynamic behaviour of the repressilator. Simulations of the repressilator show oscillations between the three inhibitory proteins, in line with the expected function of the circuit (Fig. 10g–i). In Fig. 14, we explore the behaviour of the circuit under different energy and ribosome availabilities, characterising the amplitude and period of the steady-state oscillations. The amplitude of oscillations is primarily determined by the number of ribosomes; a reduction of ribosomes from 5000 to 1000 results in an approximately 7-fold decrease in oscillation amplitude (Fig. 14a). Both energy and ribosome availability contribute to the period of oscillations. Increasing energy availability accelerates the dynamics of translation, increasing the frequency of oscillations, whereas increasing the number of ribosomes decreases the frequency of oscillations. Interestingly, similarities can be drawn with results from the toggle switch, where measures associated with the amplitude of protein changes are mostly impacted by ribosome availability, whereas temporal behaviours are determined by both energy and ribosome availability.

3.5 Parameter heterogeneity

A characteristic of cell populations is that they may experience different local environments and therefore display heterogeneity in their behaviour (Elowitz et al, 2002). Because our simple model of gene expression is computationally efficient, it is well-suited to exploring the effects of heterogeneity within cell populations. Accordingly, we explore the effects of parameter heterogeneity on the behaviour of the repressilator (Fig. 6d–f). Note that these investigations are intended to be a proof-of-concept for simulating cell populations since they assume cells operate independently of each other. In reality, cells influence one another through the transmission of extracellular signals and the competition for nutrients and resources.

We first investigate the impact of heterogeneity in energy availability on repressilator behaviour. Nakatani et al (2022) find that ATP concentrations vary substantially in populations of *E. coli*. To simulate the effect of this heterogeneity, we fit a normal distribution (mean = 3.76 mM, std = 1.64 mM) to Fig. 1A of Nakatani et al (2022). These concentrations are converted to a number of molecules, assuming a cell volume of 1 μm^3 (Weiße et al, 2015). We obtain samples of energy levels using a

truncated normal distribution with a minimum of $A = 10^4$ molecules per cell. The resulting behaviour for 10000 samples is summarised in Fig. 15a–d. Due to the relative insensitivity of the repressilator model to energy levels, the effect on the amplitude of oscillations was small, with the vast majority of samples having amplitudes between 2200–2400 (Fig. 15b). The effect on the period of oscillations was similarly small, with most samples having a period of between 60–66 mins (Fig. 15c).

Next, we consider heterogeneity in ribosome amounts. We assume that the number of free ribosomes per cell follows a similar distribution to protein counts in cells. Milo (2015) estimated that protein counts have a coefficient of variation of approximately 0.4 in bacteria. Thus, we sample ribosome counts from a log-normal distribution around a median free ribosome count of 5000 (Fig. 15e). This heterogeneity has a marked impact on the amplitudes of oscillation (∼8-fold variation; Fig. 15f), and a mild impact on the period (∼ 25% variation; Fig. 15g).

We finally explore the effect of variations in protein degradation rates. Since the protein half-life has been shown to vary 2–3 fold among populations of *E. coli* (Alber et al, 2018), we assume the protein half-life is log-normally distributed with a median half-life of 4 minutes and a 95% confidence interval spanning from $1/\sqrt{2}$ to $\sqrt{2}$ times the median. We also assume that protein half-lives are perfectly correlated between proteins, caused by global fluctuations in cell state (Elowitz et al, 2002). Using these proof-of-concept simulations, heterogeneity in protein degradation results in a wide range of variation in both the amplitudes and periods of oscillations, with approximately 2-fold variation in their values (Fig. 15i–l).

4 Discussion

In this study, we develop a new bond graph model of gene expression. In contrast to previous models in the literature (Ryall et al, 2012; Karr et al, 2012; Weiße et al, 2015; Raveh et al, 2016), our model is both simple and thermodynamically consistent, making it suitable for understanding and designing synthetic circuits in cell populations. By coupling together multiple instances of our gene expression module, we are able to reproduce behaviours of two foundational synthetic biological circuits: the toggle switch and repressilator (Gardner et al, 2000; Elowitz and Leibler, 2000). However, the approach is generalisable to other synthetic circuits involving protein synthesis, such as antithetic feedback (Briat et al, 2016) and could be used to study the effects of energy burden in such control systems. Our approach is based on bond graphs, a general framework for modelling biophysical systems. Due to their physical consistency, commonly neglected phenomena such as resource competition and retroactivity are automatically built into bond graph models (Del Vecchio, 2013; Boo et al, 2019). Additionally, their modular and hierarchical structure allows modellers to systematically couple models of synthetic circuits into larger circuits (Pan et al, 2021). Thus, we believe that bond graphs hold great potential in aiding the design of circuits in synthetic biology.

Our model of gene expression is based on a model reduction scheme to simplify the elongation reactions in translation (Sect. 2.4.2). It is common for other simple models of translation to assign rate laws based on empirical fits to data (Elowitz and Leibler,

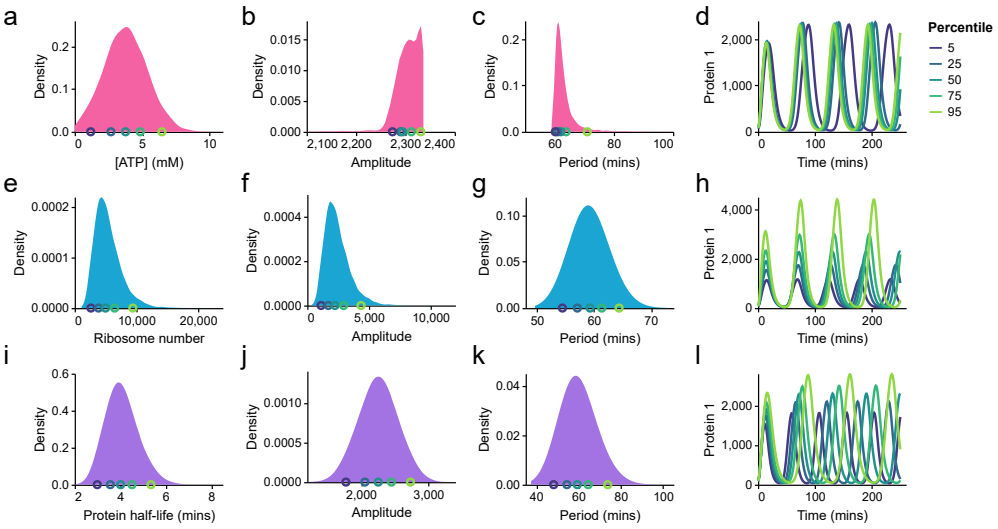

Fig. 15 Effects of parameter heterogeneity on repressilator function. Each row corresponds to distributions extracted from running 10000 simulations, with a parameter varied by sampling from a distribution. (a–d) Effects of heterogeneity in ATP concentration. Changes in ATP concentration were modelled by adjusting the parameter μ_A (14 samples were excluded from analysis as they resulted in no oscillations due to insufficient energy levels). (e–h) Effects of heterogeneity of ribosome amount, which are modelled by changing the initial ribosome number. (i–l) Effects of heterogeneity in protein decay rates, incorporated by adjusting the parameter $r_{\mathrm{deg},P}$. For each row, simulations corresponding to the 5th, 25th, 50th, 75th and 95th percentiles of the parameters are plotted on the right (d,h,l). The colours of the timecourses correspond to the circles of the same colour in the distributions.

2000; Ryall et al, 2012). By contrast, our model reduction approach is biophysically motivated, based on a detailed reaction network containing molecules and processes known to be involved in translation. As a result, our model reduction scheme preserves important constraints such as mass and energy conservation. We show that the reduced model behaves similarly to the full model at steady state and reaches the steady state on similar timescales (Figures 7–8). Furthermore, we find that the loss of dynamics in translation is unimportant when simulating synthetic gene circuits (Fig. 11). The reduced model is also computationally efficient, able to achieve runtimes hundreds of times (and possibly up to thousands of times) faster than the full model (Fig. 9). This speedup is quite significant when scaling up to large gene regulatory networks or agent-based models of large cell populations. Given that the addition of monomers to macromolecules is prevalent in biology, we believe that our model reduction scheme would also be applicable to processes such as transcription, glycogenesis and fatty acid polymerisation (Alberts, 2007).

In spite of the constraints that energy places on cells, we still have an incomplete picture of how energy constrains cellular processes and how the cell distributes its energy budget between processes (Yang et al, 2021). While previous works have studied the effects of ribosome limitation on the function of synthetic circuits (Qian et al, 2017; McBride and Del Vecchio, 2021), to the best of our knowledge, our model is the first to incorporate the known thermodynamics of ATP hydrolysis into a lumped parameter

dynamic model of translation. Our model predicts a relative insensitivity of gene expression to energy, which appears to be supported by experimental measurements (Jewett et al, 2009). Our model is also able to analyse the thermodynamic costs associated with translation. We find that the majority of energy supplied through ATP hydrolysis is dissipated during elongation rather than being stored in the peptide bonds in proteins. This inefficiency has previously been attributed to proofreading mechanisms (Hopfield, 1974); more detailed modelling of these mechanisms is required to understand performance-cost trade-offs in translation.

Our model also makes predictions about how energy and ribosome availability affect the behaviour of synthetic circuits. Interestingly, the number of free ribosomes is the primary determinant of measures associated with protein counts, whereas both energy and ribosome availability contribute to the dynamic behaviour of the circuits. We also explore the potential effects of parameter heterogeneity in cell populations. Informed by empirical measurements, we find that variations in energy levels have a limited impact on the behaviour of the repressilator (Fig. 15). However, variability in both ribosome availability and protein degradation rates result in high degrees of variation in both the amplitude and period of oscillations. Furthermore, each parameter influences different aspects of the function of the repressilator. While ribosome availability greatly influences the amplitude of oscillations, the protein degradation rate has a greater impact on the frequency of oscillations. Thus, the parameter variability that exists between cells may be responsible for cell-to-cell variability in the performance of synthetic circuits, in the absence of mechanisms of intercellular interaction (Gardner et al, 2000; Portle et al, 2009).

While we explore parameter heterogeneity as a proof-of-concept for modelling cell populations, mathematical questions remain on how to incorporate mechanisms of cell-to-cell communication. Future work will focus on incorporating our model of gene expression into an agent-based model. While the model in this study is broadly applicable to synthetic biology, the bioluminescence pathway in bacteria is a promising model system to engineer for several reasons, including (i) the ease of measuring light as the output, (ii) the involvement of quorum sensing in coordinating population-level behaviours and (iii) the use of two competing pathways for its induction (Kuttler and Hense, 2008). Furthermore, the multi-physical nature of bond graphs allows them to naturally account for the conversion of chemical energy to light (Gawthrop and Pan, 2022). Modelling the intracellular regulatory mechanisms associated with quorum sensing alongside intercellular communication has the potential to enable the engineering of synchronised behaviour of cell populations (Prindle et al, 2012; Wu et al, 2020).

The work in this study contributes towards the unification of modelling metabolism and gene regulation. The link between energy and enzymes is fundamental to much of biology, but has been underexplored, despite the fact that these two subsystems have been extensively studied individually (Noor et al, 2016; Klosik et al, 2017). While metabolism has previously been modelled using bond graphs (Gawthrop et al, 2021), this paper presents the first bond graph model of gene expression. Using the modular structure of bond graphs, it is relatively straightforward to couple metabolic processes to gene regulation. Furthermore, the level of detail in the metabolic processes being

modelled can be adjusted, from simple models (Weiße et al, 2015) to large-scale models (Stanford et al, 2013). Such models would allow us to better understand how cells distribute energy between metabolic, signalling and gene regulatory pathways, a problem central to much of systems and synthetic biology (Tran et al, 2015; Weiße et al, 2015; Karr et al, 2012).

Due to the simple nature of the model, certain details of gene regulation have been omitted. While we assume that elongation involves a single reaction per ATP molecule, tRNA is in reality charged by amino acids through elongation factors, which involves a more complex series of reactions (Hu et al, 2020). The elongation rate is dependent on tRNA availability, and therefore specific to each codon. We also lump energy into a single molecule, but in reality, it is composed of the hydrolysis of ATP into ADP, Pi and H^+. Accounting for such details will be the subject of future work.

Acknowledgments. S.T.J. is supported by the Australian Research Council (Project No. DE200100988). P.J.G. would like to thank the Faculty of Engineering and Information Technology, University of Melbourne, for its support via a Professorial Fellowship. M.P. was supported by a Postdoctoral Research Fellowship from the School of Mathematics and Statistics at the University of Melbourne.

Data availability. The bond graph models were simulated in Julia using the BondGraphs.jl package (Forrest et al, 2023). The code is available on GitHub at https://github.com/mic-pan/bg_gene_expression.

Declarations

Competing Interests. None declared.

A Parameters

This section discusses how parameter values are chosen. Table 1 summarises the default values, with the justification discussed in later sections.

A.1 Molecule counts

Energy: Under nutrient-rich conditions, the ATP concentration has been estimated to be 9.63 mM in *E. coli* (Park et al, 2016). Assuming that the volume of an *E. coli* cell is 1 μm^3 (Weiße et al, 2015), this corresponds to 5.8×10^6 molecules of ATP per cell. We use this as the reference number of energy molecules in a cell.

mRNA: In simulating translation in isolation, we assume there are 10 mRNA molecules per cell. This corresponds to a relatively high transcript count for a single gene (Milo, 2015).

Ribosomes: The number of ribosomes in an *E. coli* cell can vary between 6800–72000 (Milo, 2015). However, the majority of ribosomes are occupied in translating proteins. At high growth rates, Dai et al (2016) estimated that 80% of ribosomes are actively translating protein. Thus, we assume there were 5000 free ribosomes per cell, corresponding to a total ribosome number of approximately 25000. Models are

Table 2 Default values used in gene expression model.

	Description	Default value	Section	References
x_A	Number of energy molecules	5.8×10^6 molecs	A.1	Park et al (2016); Weiße et al (2015)
x_M	Number of mRNA molecules	10 molecs	A.1	Milo (2015)
x_R	Number of ribosomes	5000 molecs	A.1	Milo (2015); Dai et al (2016)
n	ATP molecules required to synthesise protein	1200 molecs	A.2	Weiße et al (2015)
K_M	Thermodynamic parameter for mRNA	1 molecs^{-1}	A.2	–
K_R	Thermodynamic parameter for free ribosome	1 molecs^{-1}	A.2	–
K_A	Thermodynamic parameter for energy	83.65 molecs^{-1}	A.2	Song and Hyeon (2020); Park et al (2016)
K_{C_0}	Thermodynamic parameter for initial ribosomal complex	1 molecs^{-1}	A.2	–
α	Ratio of parameters between ribosomal complexes	3.49	A.2	Song and Hyeon (2020)
K_{C_1}	Thermodynamic parameter for ribosomal complex 1	α molecs^{-1}	A.2	–
K_{C_i}	Thermodynamic parameter for ribosomal complexes ($i > 1$)	α^i molecs^{-1}	A.2	–
K_P	Thermodynamic parameter for protein	$\alpha^n / e^{\bar{\mu}_{\text{termination}}}$ molecs^{-1}	A.2	Zeldovich et al (2007)
$\bar{r}_{\text{Tc}}$	Transcription rate	1.13×10^{-2} molecs/min	A.3.1	Weiße et al (2015)
R_{bA}	Substrate binding parameter for transcription	366.4	A.3.1	Weiße et al (2015)
R_{bM}	Product binding parameter for transcription	10^6	A.3.1	–
R_{bI}	Inhibition parameter for transcription	$30K_P$	A.3.1	Weiße et al (2015)
r_b	Translation initiation rate	0.001 molecs/min	A.3.2	–
r_1	Rate parameter for first elongation reaction	1.04 molecs/min	A.3.2	Bremer et al (1996)
r_i	Rate parameters for ith elongation reaction ($i > 1$)	$5040/(e^{\bar{\mu}_A} K_{C_i})$ molecs/min	A.3.2	Bremer et al (1996)
r_{el}	Rate parameter for elongation (reduced model)	1.04×10^{-5} molecs/min	A.3.2	Bremer et al (1996)
r_t	Translation termination rate	$504000/K_{C_n}$ molecs/min	A.3.2	–
$r_{\text{deg},M}$	mRNA degradation rate	0.1 molecs/min	A.3.3	Weiße et al (2015)
$r_{\text{deg},P}$	Protein degradation rate	$0.02/K_P$ molecs/min	A.3.3	Weiße et al (2015)
$\bar{\mu}_{X,M}$	Chemical potential of mRNA degradation products	$\ln(10^{-6})$	A.3.3	–
$\bar{\mu}_{X,P}$	Chemical potential of protein degradation products	0	A.3.3	–

initialised with virtually all ribosomes in the free state; complex amounts are initialised to 10^{-6} molecules to avoid infinite potentials.

A.2 Thermodynamic parameters

Reference species: In the absence of free energy measurements, the mRNA (M) and ribosome (R) molecules are assumed to be reference species, so that their thermodynamic parameters are set to 1.

Energy: We assume that under the reference condition, the free energy of ATP hydrolysis is $-20k_bT$, corresponding to a Gibbs free energy of 49.5 kJ/mol at a temperature of 298 K. With a potential of $\bar{\mu}_A = 20$ and molecular count of $x_A = 5.8 \times 10^6$ molecules, the thermodynamic parameter is $K_A = 83.65$ molecules^{-1} (Eq. (1)).

Our representation of ATP hydrolysis is a simplified version of the true biological reaction $\mathrm{ATP} + \mathrm{H_2O} \rightleftharpoons \mathrm{ADP} + \mathrm{Pi} + \mathrm{H^+}$. Thus, our model only considers ATP and neglects the other molecules. However, under conditions where the concentrations of the other molecules are constant, it can be shown that the equations reduce to our model.

Complexes: From Weiße et al (2015), the dissociation constant of the translation initiation reaction ($M + R \rightleftharpoons C_0$) is $K_b^d = K_{C_0}/(K_M K_R) = 1$. Then, since $K_M = K_R = 1$, it follows that $K_{C_0} = 1$.

We assume that the standard potentials of the chemical complexes differ by the energy required to form a peptide bond, or approximately $5k_BT$ of energy (Song and Hyeon, 2020). Since four ATP molecules are required to form a peptide bond, the standard potentials of successive complexes differ by $5/4k_BT$ and hence $\alpha = K_{C_{i+1}}/K_{C_i} = e^{5/4} = 3.49$. This corresponds to $\hat{A} = e^{\bar{\mu}_A - 5/4} = 1.38 \times 10^8$ under reference conditions (Eq. (19)).

Protein: We assume that the termination of translation involves free energy contributions from protein folding and the dissociation of the protein synthesis complex. Because the free energy of protein folding ($\bar{\mu}_{\text{folding}} = -8.5$; Zeldovich et al (2007)) alone is insufficient to result in approximate irreversibility of termination, we assume a free energy of dissociation of $\bar{\mu}_{\text{dissociation}} = -11.5$ to result in a standard free energy of $\mu_{\text{termination}} = \bar{\mu}_{\text{folding}} + \bar{\mu}_{\text{dissociation}} = -20$. Hence, the thermodynamic parameter is $K_P = K_{C_n} e^{\bar{\mu}_{\text{termination}}} = \alpha^n e^{-20}$.

Protein length: We assume $n = 1200$, which corresponds to a protein length of 300 amino acids (Weiße et al, 2015).

Adjustment for mRNA copy number: When simulating simulation in isolation, an adjustment is made to the above parameters. In considering the reverse reaction of $C_0 + A \underset{k_{r1}}{\overset{k_{f1}}{\rightleftharpoons}} C_1 + M$, a complex may only reoccupy the first site of its own transcript. To account for this, the reverse rate parameter k_{r1} must be scaled down by the total number of mRNA molecules M_{tot}, i.e. $k_{r1} = k_r/M_{\text{tot}}$. Thus, $K_{C_1} = \alpha/(K_M M_{\text{tot}})$, $K_{C_i} = \alpha^i/(K_M M_{\text{tot}})$, $K_P = \alpha^n/(K_M M_{\text{tot}})$. However, since the number of mRNA molecules becomes variable once transcription and translation are coupled together, this adjustment is not applied under the coupled model.

A.3 Rate parameters

A.3.1 Transcription

Transcription parameters are calculated by matching enzyme kinetics behaviour to that of Weiße et al (2015). The transcription rate law follows a Michaelis-Menten equation $v_{\mathrm{Tc}} = wA/(A+\theta)$ with maximum flux $w = 4.14$ molecules/min and a half-saturation constant of $\theta = 4.38$ molecules. By comparing to Eq. (23),

$$\begin{aligned} w/\theta &= \bar{r}_{\mathrm{Tc}} K_A, \\ \theta &= R_{bA}/K_A. \end{aligned}$$

Solving the above equations, $\bar{r}_{\mathrm{Tc}} = 1.13 \times 10^{-2}$ molecs/min and $R_{bA} = 366.4$. We assume $R_{bP} = 10^6$ so that the reverse reaction only saturates at high concentrations, and that $R_{bI} = 30e^{\bar{\mu}_I^0}$ to achieve inhibition at 30 molecules of inhibitor.

A.3.2 Translation

Binding rate: The binding rate constant is set to $r_b = 0.001$ to achieve biologically feasible protein copy numbers (1000's to 10000's) at steady-states with 10–100 mRNA molecules per cell (Milo, 2015).

Elongation: Under nutrient-rich conditions, the elongation rate has been measured to be 1260 amino acids per minute (Bremer et al, 1996). Since we model four elongation steps per amino acid, the forward elongation rate is $k_f A = 4 \times 1260 = 5040$ molecules/min. Then, since $k_f = r_i K_A K_{C_{i-1}}$ for $i = 1, \ldots, n$, $r_i = 5040/(K_A A)/K_{C_i} = 5040/(e^{20} K_{C_i})$. In the reduced model, these rate parameters correspond to $r_{\mathrm{el}} = k_f A\alpha/e^{20}/K_{C_1} = 1.04 \times 10^{-5}$ molecules/min.

Termination: Termination is assumed to be fast relative to elongation, with a forward rate constant 100 times that of an elongation step. Thus, $r_t = 100 k_f A/K_{C_n} = 504000/K_{C_n}$.

A.3.3 Degradation

mRNA degradation: The mRNA degradation decay rate is set to 0.1 min^{-1}, in line with Weiße et al (2015). The reaction rate associated with degradation is therefore set to $r_{\mathrm{deg,M}} = 0.1/K_M = 0.1$ molecules/min. A negative degradation product potential of $\bar{\mu}_{X,M} = \ln(10^{-6})$ is used to approximate an irreversible reaction.

Protein degradation: The protein degradation decay rate is set to 0.02 min^{-1}, so that proteins degrade slower than mRNA. The reaction rate associated with degradation is therefore set to $r_{\mathrm{deg,P}} = 0.02/K_P$ molecules/s. A degradation product potential of $\bar{\mu}_{X,P} = 0$ is used to approximate an irreversible reaction.

A.4 Toggle switch and repressilator

The following parameters are modified in simulating the toggle switch and repressilator:

- To increase expression levels, the binding constant is increased to $r_b = 0.01$ molecules/min
- The inhibition parameter is set to $R_I = (100K_P)^h$ so that inhibition occurs at protein concentrations on the order of 100 molecules Weiße et al (2015)
- The mRNA degradation rate is increased to $\log(2)/2$ min^{-1}, consistent with Weiße et al (2015)
- The protein degradation rate is increased to $\log(2)/4$ min^{-1}, consistent with Weiße et al (2015)

A.5 Simplification under excess energy

A.5.1 Gene expression

Here we consider the reduced gene expression model with one lumped complex (Sect. 2.5). Under conditions where A is high, Eq. (23) becomes

$$v_{\mathrm{Tc}} \approx v^*_{\mathrm{Tc}} = \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + e^{\bar{\mu}_I}/R_{bI}}.$$

Also, due to the high amount of A, the elongation reactions can be assumed to be irreversible and instantaneous, so complexes are immediately moved to their terminal complex. Thus, the reaction network reduces to

$$\begin{aligned} \varnothing &\to M \to \varnothing \\ M + R &\to M + C_n \\ C_n &\rightleftharpoons R + P \\ P &\to \varnothing. \end{aligned}$$

For ease of analysis, we make the additional assumptions that (i) the mRNA transcription and degradation dynamics are faster than translation and (ii) termination is instantaneous and irreversible. Under these further assumptions, the reaction network becomes

$$\begin{aligned} \varnothing &\to M \to \varnothing \\ M + R &\to M + R + P \\ P &\to \varnothing, \end{aligned}$$

where the mRNA dynamics follows the equation

$$\begin{aligned} \frac{\mathrm{d}x_M}{\mathrm{d}t} &\approx v^*_{\mathrm{Tc}}(\bar{\mu}_I) - r_{\mathrm{deg},M} K_M x_M = 0 \\ \Rightarrow x_M &= \frac{v^*_{\mathrm{Tc}}(\bar{\mu}_I)}{r_{\mathrm{deg},M} K_M}. \end{aligned}$$

Using the steady-state value of M, the equation for protein dynamics becomes

$$\frac{\mathrm{d}x_P}{\mathrm{d}t} = r_b K_M K_R x_M x_R - r_{\mathrm{deg},P} K_P x_P = r_b K_M K_R x_R \frac{v^*_{\mathrm{Tc}}(\bar{\mu}_I)}{r_{\mathrm{deg},M} K_M} - r_{\mathrm{deg},P} K_P x_P.$$

Thus, under the limit of high energy levels, the governing differential equations are of a similar form to previous works (Gardner et al, 2000).

At steady state,

$$x_P = \frac{r_b K_M K_R x_R v^*_{\mathrm{Tc}}(\bar{\mu}_I)}{r_{\mathrm{deg},M} K_M r_{\mathrm{deg},P} K_P},$$

and the translation rate is the positive term of $\mathrm{d}x_P/\mathrm{d}t$:

$$v_{\mathrm{trans}} = \frac{r_b K_M K_R x_R v^*_{\mathrm{Tc}}(\bar{\mu}_I)}{r_{\mathrm{deg},M} K_M}.$$

A.5.2 Toggle switch

Under the above assumptions, the toggle switch follows the following equations at steady state:

$$\frac{\mathrm{d}x_{P1}}{\mathrm{d}t} = \frac{r_b K_M K_R x_R}{r_{\mathrm{deg},M} K_M} \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + (K_P x_{P2})^2 / R_{bI}} - r_{\mathrm{deg},P} K_P x_{P1},$$
$$\frac{\mathrm{d}x_{P2}}{\mathrm{d}t} = \frac{r_b K_M K_R x_R}{r_{\mathrm{deg},M} K_M} \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + (K_P x_{P1})^2 / R_{bI}} - r_{\mathrm{deg},P} K_P x_{P2}.$$

A phase plane of this system is shown in Fig. 16, with the nullclines

$$x_{P1} = x_{P1}(x_{P2}) = \frac{r_b K_M K_R x_R}{r_{\mathrm{deg},M} K_M r_{\mathrm{deg},P} K_P} \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + (K_P x_{P2})^2 / R_{bI}},$$
$$x_{P2} = x_{P2}(x_{P1}) = \frac{r_b K_M K_R x_R}{r_{\mathrm{deg},M} K_M r_{\mathrm{deg},P} K_P} \frac{\bar{r}_{\mathrm{Tc}} R_{bA}}{1 + (K_P x_{P1})^2 / R_{bI}}.$$

Thus, under the reference parameters, there are three fixed points, with two stable states. As the number of ribosomes decreases, the stable points move towards the origin, leading to a decrease in bistability index.

A.6 Numerics

For some species, the thermodynamic parameter K is larger than the maximum value permitted by 64-bit floats. However, these large parameters are counteracted by small reaction rates r to result in moderate fluxes. To ensure that values remain numerically defined, some components use log-transformed parameters. In particular, the log-transformed species is defined by the equation

$$\bar{\mu} = \bar{\mu}^0 + \ln(x),$$

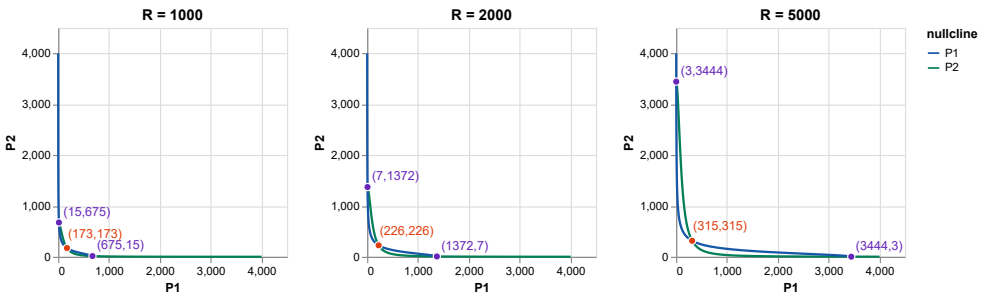

Fig. 16 A phase plane of the toggle switch. Lines show the nullclines of the system. Purple points represent the stable fixed points, and the orange points are the unstable fixed points.

where $\bar{\mu}^0 = \ln(K)$; and log-transformed reactions are defined by the equation

$$v = e^{\bar{\mu}^a + \bar{\mu}^f} - e^{\bar{\mu}^a + \bar{\mu}^r},$$

where $\bar{\mu}^a = \ln(\kappa)$. Regulatory parameters such as R_{bI} in Eq. (9) can be log-transformed in a similar manner. This formulation is mathematically equivalent to the standard thermodynamic parameters, but avoids the need to evaluate large numbers.

B Supplementary figures

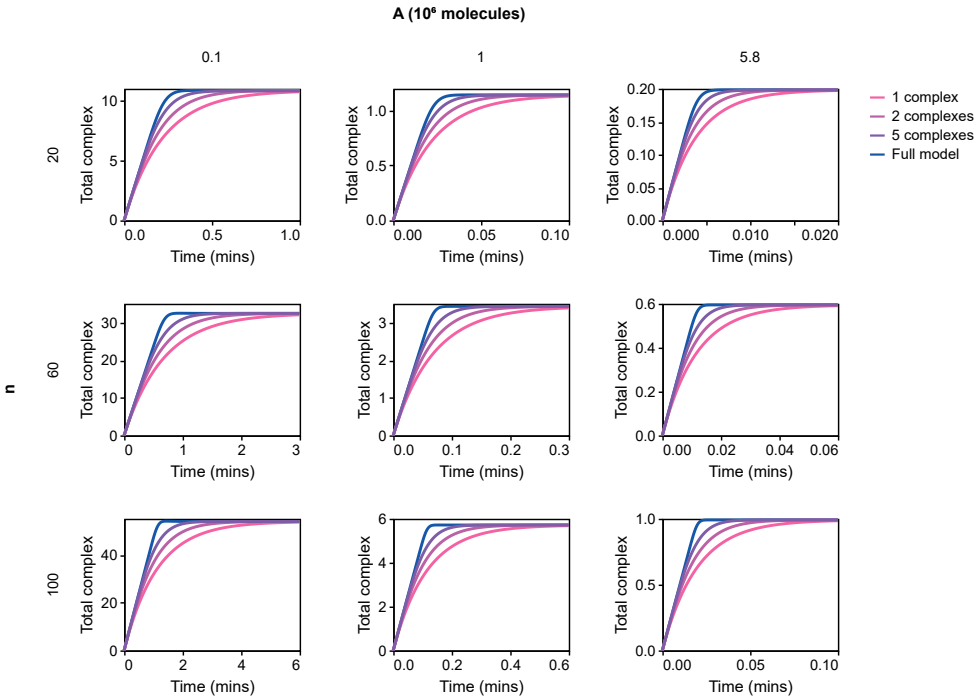

Fig. 17 Comparison of total complex amount for the full model. The energetic state A is varied horizontally, and the chain length n is varied vertically. The translation rate is given in proteins per minute. Simulations are run with 10 mRNA molecules, 5000 ribosomes and $\bar{\mu_P} = -15$ to approximate the immediate removal of proteins.

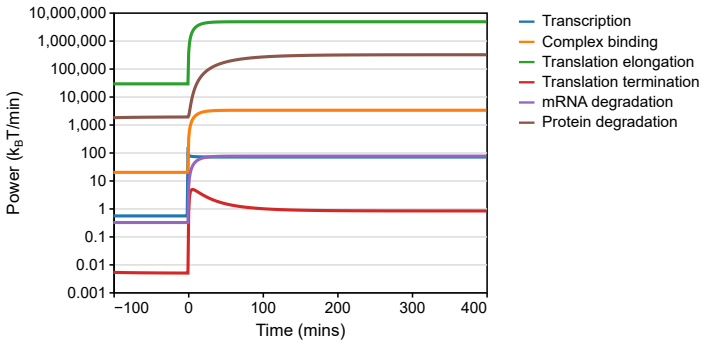

Fig. 18 Power consumption of gene regulation. The gene expression model in Fig. 10a–c is simulated, and the power dissipation rate of each process is plotted against time. The power dissipation rate of each reaction is defined as the sum of power through all ports of the corresponding reaction. The power disspation for elongation is defined as the sum of power dissipation rates for the first elongation reaction (Re:r_1) and the elongation module $\mathrm{Re}_{\mathrm{tl}}$.

References

Alber AB, Paquet ER, Biserni M, et al (2018) Single live cell monitoring of protein turnover reveals intercellular variability and cell-cycle dependence of degradation rates. Molecular Cell 71(6):1079–1091.e9. https://doi.org/10.1016/j.molcel.2018.07.023

Alberts B (2007) Molecular Biology of the Cell, 5th edn. Garland Science, New York

Atkins P, de Paula J (2006) Physical Chemistry for the Life Sciences. Oxford University Press, Oxford, England, UK

Basu S, Gerchman Y, Collins CH, et al (2005) A synthetic multicellular system for programmed pattern formation. Nature 434:1130–1134. https://doi.org/10.1038/nature03461

Boo A, Ellis T, Stan GB (2019) Host-aware synthetic biology. Current Opinion in Systems Biology 14:66–72. https://doi.org/10.1016/j.coisb.2019.03.001

Boo A, Ledesma Amaro R, Stan GB (2021) Quorum sensing in synthetic biology: A review. Current Opinion in Systems Biology 28:100378. https://doi.org/10.1016/j.coisb.2021.100378

Bremer H, Dennis PP, et al (1996) Modulation of chemical composition and other parameters of the cell by growth rate. Escherichia coli and Salmonella: cellular and molecular biology 2(2):1553–69

Briat C, Gupta A, Khammash M (2016) Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks. Cell Systems 2(1):15–26. https://doi.org/10.1016/j.cels.2016.01.004

Caringella G, Bandiera L, Menolascina F (2023) Recent advances, opportunities and challenges in cybergenetic identification and control of biomolecular networks. Current Opinion in Biotechnology 80:102893. https://doi.org/10.1016/j.copbio.2023.102893

Chen JR (2006) Improving the speed of the genetic toggle switch without sacrificing its dynamic stability. Physical Review E 73(4):041901. https://doi.org/10.1103/PhysRevE.73.041901

Church GM, Elowitz MB, Smolke CD, et al (2014) Realizing the potential of synthetic biology. Nature Reviews Molecular Cell Biology 15(4):289–294. https://doi.org/10.1038/nrm3767

Dai X, Zhu M, Warren M, et al (2016) Reduction of translating ribosomes enables escherichia coli to maintain elongation rates during slow growth. Nature Microbiology 2(16231):1–9. https://doi.org/10.1038/nmicrobiol.2016.231

Del Vecchio D (2013) A control theoretic framework for modular analysis and design of biomolecular networks. Annual Reviews in Control 37(2):333–345. https://doi.org/10.1016/j.arcontrol.2013.09.011

Dowling CV, Cevaal PM, Faria M, et al (2022) On predicting heterogeneity in nanoparticle dosage. Mathematical Biosciences 354:108928. https://doi.org/10.1016/j.mbs.2022.108928

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403(6767):335–338. https://doi.org/10.1038/35002125

Elowitz MB, Levine AJ, Siggia ED, et al (2002) Stochastic gene expression in a single cell. Science 297(5584):1183–1186. https://doi.org/10.1126/science.1070919

Forrest J, Rajagopal V, Stumpf MPH, et al (2023) BondGraphs.jl: composable energy-based modelling in systems biology. Bioinformatics 39(10):btad578. https://doi.org/10.1093/bioinformatics/btad578

Gander MW, Vrana JD, Voje WE, et al (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. Nature Communications 8(15459):1–11. https://doi.org/10.1038/ncomms15459

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. Nature 403(6767):339–342. https://doi.org/10.1038/35002131

Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. Nature Reviews Molecular Cell Biology 8:113–126. https://doi.org/10.1038/nrm2104

Gawthrop PJ (2021) Energy-based modeling of the feedback control of biomolecular systems with cyclic flow modulation. IEEE Transactions on NanoBioscience 20(2):183–192. https://doi.org/10.1109/TNB.2021.3058440

Gawthrop PJ, Crampin EJ (2014) Energy-based analysis of biochemical cycles using bond graphs. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 470(2171):20140459. https://doi.org/10.1098/rspa.2014.0459

Gawthrop PJ, Pan M (2022) Network thermodynamics of biological systems: A bond graph approach. Mathematical Biosciences 352:108899. https://doi.org/10.1016/j.mbs.2022.108899

Gawthrop PJ, Pan M, Crampin EJ (2021) Modular dynamic biomolecular modelling with bond graphs: the unification of stoichiometry, thermodynamics, kinetics and data. Journal of the Royal Society Interface 18(181):20210478. https://doi.org/10.1098/rsif.2021.0478

Gorochowski TE (2016) Agent-based modelling in synthetic biology. Essays in Biochemistry 60(4):325–336. https://doi.org/10.1042/EBC20160037

Gray P, Meek S, Griffiths P, et al (2018) Synthetic biology in australia: An outlook to 2030. Tech. Rep. 978-0-6483303-1-8, Australian Council of Learned Academies, Melbourne, Victoria, Australia

Hindmarsh AC, Brown PN, Grant KE, et al (2005) Sundials: Suite of nonlinear and differential/algebraic equation solvers. ACM Transactions on Mathematical Software 31(3):363–396. https://doi.org/10.1145/1089014.1089020

Hopfield JJ (1974) Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. Proceedings of the National Academy of Sciences 71(10):4135–4139. https://doi.org/10.1073/pnas.71.10.4135

Hu XP, Dourado H, Schubert P, et al (2020) The protein translation machinery is expressed for maximal efficiency in Escherichia coli. Nature Communications 11(5260):1–10. https://doi.org/10.1038/s41467-020-18948-x

Hurtado PJ, Kirosingh AS (2019) Generalizations of the 'Linear Chain Trick': incorporating more flexible dwell time distributions into mean field ODE models. Journal of Mathematical Biology 79(5):1831–1883. https://doi.org/10.1007/s00285-019-01412-w

Jacques N, Dreyfus M (1990) Translation initiation in Escherichia coli: old and new questions. Molecular Microbiology 4(7):1063–1067. https://doi.org/10.1111/j.1365-2958.1990.tb00679.x

Jewett MC, Miller ML, Chen Y, et al (2009) Continued protein synthesis at low [ATP] and [GTP] enables cell adaptation during energy limitation. Journal of Bacteriology URL https://journals.asm.org/doi/10.1128/JB.00852-08

Karr JR, Sanghvi JC, Macklin DN, et al (2012) A whole-cell computational model predicts phenotype from genotype. Cell 150(2):389–401. https://doi.org/10.1016/j.cell.2012.05.044

Keener J, Sneyd J (2009) Mathematical Physiology. Springer, New York, URL http://link.springer.com/10.1007/978-0-387-75847-3

Kempes CP, Wolpert D, Cohen Z, et al (2017) The thermodynamic efficiency of computations made in cells across the range of life. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 375(2109):20160343. https://doi.org/10.1098/rsta.2016.0343

Klosik DF, Grimbs A, Bornholdt S, et al (2017) The interdependent network of gene regulation and metabolism is robust where it needs to be. Nature Communications 8(534):1–9. https://doi.org/10.1038/s41467-017-00587-4

Kuttler C, Hense BA (2008) Interplay of two quorum sensing regulation systems of vibrio fischeri. Journal of Theoretical Biology 251(1):167–180. https://doi.org/10.1016/j.jtbi.2007.11.015

López-Otín C, Overall CM (2002) Protease degradomics: A new challenge for proteomics. Nature Reviews Molecular Cell Biology 3:509–519. https://doi.org/10.1038/nrm858

Machado D, Costa RS, Rocha M, et al (2011) Modeling formalisms in systems biology. AMB Express 1(1):45. https://doi.org/10.1186/2191-0855-1-45

Macklin DN, Ahn-Horst TA, Choi H, et al (2020) Simultaneous cross-evaluation of heterogeneous E. coli datasets via mechanistic simulation. Science 369(6502). URL https://www.science.org/doi/10.1126/science.aav3751

McBride CD, Del Vecchio D (2021) Predicting composition of genetic circuits with resource competition: Demand and sensitivity. ACS Synthetic Biology 10(12):3330–3342. https://doi.org/10.1021/acssynbio.1c00281

Milo P (2015) Cell Biology by the Numbers. Taylor & Francis, Andover, England, UK, https://doi.org/10.1201/9780429258770

Nagarajan K, Ni C, Lu T (2022) Agent-based modeling of microbial communities. ACS Synthetic Biology 11(11):3564–3574. https://doi.org/10.1021/acssynbio.2c00411

Nakatani RJ, Itabashi M, Yamada TG, et al (2022) Intercellular interaction mechanisms promote diversity in intracellular ATP concentration in Escherichia

coli populations. Scientific Reports 12(17946):1–11. https://doi.org/10.1038/s41598-022-22189-x

Noor E, Flamholz A, Bar-Even A, et al (2016) The protein cost of metabolic fluxes: Prediction from enzymatic rate laws and cost minimization. PLOS Computational Biology 12(11):e1005167. https://doi.org/10.1371/journal.pcbi.1005167

Pan M, Gawthrop PJ, Tran K, et al (2018) Bond graph modelling of the cardiac action potential: implications for drift and non-unique steady states. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 474(2214):20180106. https://doi.org/10.1098/rspa.2018.0106

Pan M, Gawthrop PJ, Tran K, et al (2019) A thermodynamic framework for modelling membrane transporters. Journal of Theoretical Biology 481:10–23. https://doi.org/10.1016/j.jtbi.2018.09.034

Pan M, Gawthrop PJ, Cursons J, et al (2021) Modular assembly of dynamic models in systems biology. PLOS Computational Biology 17(10):e1009513. https://doi.org/10.1371/journal.pcbi.1009513

Park JO, Rubin SA, Xu YF, et al (2016) Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nature Chemical Biology 12:482–489. https://doi.org/10.1038/nchembio.2077

Polettini M, Esposito M (2014) Irreversible thermodynamics of open chemical networks. i. emergent cycles and broken conservation laws. Journal of Chemical Physics 141(2):024117. https://doi.org/10.1063/1.4886396

Portle S, Iadevaia S, San KY, et al (2009) Environmentally-modulated changes in fluorescence distribution in cells with oscillatory genetic network dynamics. Journal of Biotechnology 140(3):203–217. https://doi.org/10.1016/j.jbiotec.2009.01.011

Prabhakar A, Pavlov MY, Zhang J, et al (2023) Dynamics of release factor recycling during translation termination in bacteria. Nucleic Acids Research 51(11):5774–5790. https://doi.org/10.1093/nar/gkad286

Prindle A, Samayoa P, Razinkov I, et al (2012) A sensing array of radically coupled genetic 'biopixels'. Nature 481:39–44. https://doi.org/10.1038/nature10722

Qian Y, Huang HH, Jiménez JI, et al (2017) Resource competition shapes the response of genetic circuits. ACS Synthetic Biology 6(7):1263–1272. https://doi.org/10.1021/acssynbio.6b00361

Raveh A, Margaliot M, Sontag ED, et al (2016) A model for competition for ribosomes in the cell. Journal of The Royal Society Interface 13(116):20151062. https://doi.org/10.1098/rsif.2015.1062

Ryall KA, Holland DO, Delaney KA, et al (2012) Network reconstruction and systems analysis of cardiac myocyte hypertrophy signaling. Journal of Biological Chemistry 287(50):42259–42268. https://doi.org/10.1074/jbc.M112.382937

Salis HM (2011) The ribosome binding site calculator. In: Methods in Enzymology, vol 498. Academic Press, Cambridge, MA, USA, p 19–42, https://doi.org/10.1016/B978-0-12-385120-8.00002-4

Shallom D, Naiger D, Weiss S, et al (2021) Accelerating whole-cell simulations of mRNA translation using a dedicated hardware. ACS Synthetic Biology 10(12):3489–3506. https://doi.org/10.1021/acssynbio.1c00415

Shaw WM, Yamauchi H, Mead J, et al (2019) Engineering a model cell for rational tuning of GPCR signaling. Cell 177(3):782–796. https://doi.org/10.1016/j.cell.2019.02.023

Slusarczyk AL, Lin A, Weiss R (2012) Foundations for the design and implementation of synthetic genetic circuits. Nature Reviews Genetics 13:406–420. https://doi.org/10.1038/nrg3227

Song Y, Hyeon C (2020) Thermodynamic cost, speed, fluctuations, and error reduction of biological copy machines. Journal of Physical Chemistry Letters 11(8):3136–3143. https://doi.org/10.1021/acs.jpclett.0c00545

Stanford NJ, Lubitz T, Smallbone K, et al (2013) Systematic construction of kinetic models from genome-scale metabolic networks. PLoS ONE 8(11):e79195. https://doi.org/10.1371/journal.pone.0079195

Tran K, Loiselle DS, Crampin EJ (2015) Regulation of cardiac cellular bioenergetics: mechanisms and consequences. Physiol Rep 3(7):e12464. https://doi.org/10.14814/phy2.12464

Turberville A, Semple H, Davies G, et al (2022) A perspective on the discovery of enzyme activators. SLAS Discovery 27(8):419–427. https://doi.org/10.1016/j.slasd.2022.09.001

Vind J, Sørensen MA, Rasmussen MD, et al (1993) 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 231(3):678–688. https://doi.org/10.1006/jmbi.1993.1319

Weiße AY, Oyarzún DA, Danos V, et al (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. Proceedings of the National Academy of Sciences 112(9):E1038–E1047. https://doi.org/10.1073/pnas.1416533112

Wu S, Liu J, Liu C, et al (2020) Quorum sensing for population-level control of bacteria and potential therapeutic applications. Cellular and Molecular Life Sciences

77(7):1319–1343. https://doi.org/10.1007/s00018-019-03326-8

Yang X, Heinemann M, Howard J, et al (2021) Physical bioenergetics: Energy fluxes, budgets, and constraints in cells. Proceedings of the National Academy of Sciences 118(26):e2026786118. https://doi.org/10.1073/pnas.2026786118

Zeldovich KB, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. Proceedings of the National Academy of Sciences 104(41):16152–16157. https://doi.org/10.1073/pnas.0705366104