ACS Synthetic Biology

Research Article

pubs.acs.org/synthbio

A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes

Thomas E. Gorochowski,*[,†,‡] Irem Avcilar-Kucukgoze,[§] Roel A. L. Bovenberg,[†,∥] Johannes A. Roubos,[†] and Zoya Ignatova[§,⊥]

[†]DSM Biotechnology Center, P.O. Box 1, 2600 MA Delft, The Netherlands

[‡]BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, U.K.

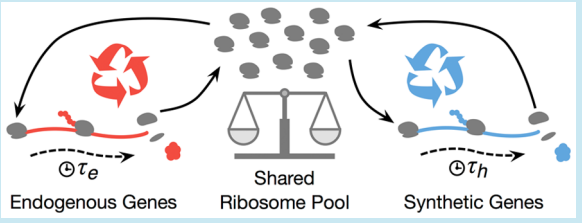
[§]Biochemistry, Institute of Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany

[∥]Synthetic Biology and Cell Engineering, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, The Netherlands

[⊥]Biochemistry and Molecular Biology, Department of Chemistry, University of Hamburg, 20146 Hamburg, Germany

Supporting Information

ABSTRACT: Cells contain a finite set of resources that must be distributed across many processes to ensure survival. Among them, the largest proportion of cellular resources is dedicated to protein translation. Synthetic biology often exploits these resources in executing orthogonal genetic circuits, yet the burden this places on the cell is rarely considered. Here, we develop a minimal model of ribosome allocation dynamics capturing the demands on translation when expressing a synthetic construct together with endogenous genes required for the maintenance of cell physiology. Critically, it contains three key variables related to design parameters of the synthetic construct covering transcript abundance, translation initiation rate, and elongation time. We show that model-predicted changes in ribosome allocation closely match experimental shifts in synthetic protein expression rate and cellular growth. Intriguingly, the model is also able to accurately infer transcript levels and translation times after further exposure to additional ambient stress. Our results demonstrate that a simple model of resource allocation faithfully captures the redistribution of protein synthesis resources when faced with the burden of synthetic gene expression and environmental stress. The tractable nature of the model makes it a versatile tool for exploring the guiding principles of efficient heterologous expression and the indirect interactions that can arise between synthetic circuits and their host chassis because of competition for shared translational resources.



KEYWORDS: *protein biosynthesis, translation, synthetic biology, systems biology*

The maintenance of balanced gene expression and energy metabolism is essential for cellular growth and viability. Exposure to adverse environmental conditions can perturb this optimal state, causing cells to adapt through dynamic reprogramming of internal activities.[1−3] The growing use of microorganisms as production chassis for synthetic protein products and the need for them to function at large scales where environmental conditions cannot always be optimally maintained[4−6] mean that cells are increasingly facing simultaneous production demands and environmental stress that can impact their normal physiology and impair productivity. To ensure optimal production yields, it is crucial to understand the deviations in cellular behaviors and address the underlying mechanisms. Most studies have focused on descriptions of the many steps involved in protein synthesis that are not scalable and contain many unknown parameters.[7−9] Furthermore, the combined effect of synthetic gene expression burden and ambient stress remains largely unexplored.

A significant proportion of cellular resources are dedicated to protein synthesis, with approximately 50% of the energy in bacterial cells used for translation.[10] This places translation among the central processes that regulate the response of the cell to various types of stress, allowing the rapid diversion of resources to the maintenance of cell viability.[11] The critical role of ribosomes in supporting general growth has also been empirically shown,[12−14] highlighting translation as a major controlling hub of cellular physiology. Efforts have been made to model the dynamics of ribosome redistribution by overexpressing foreign proteins to place varying burdens on this shared resource.[9] Generally, these studies have taken a mechanistic view of the discrete steps of translation, e.g., initiation, the multiple successive translocations of the ribosome from codon to codon along a transcript, and termination.[7,8] While capable of reproducing some features of the cellular response to differing expression demands, the complexity and

Received: February 1, 2016

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

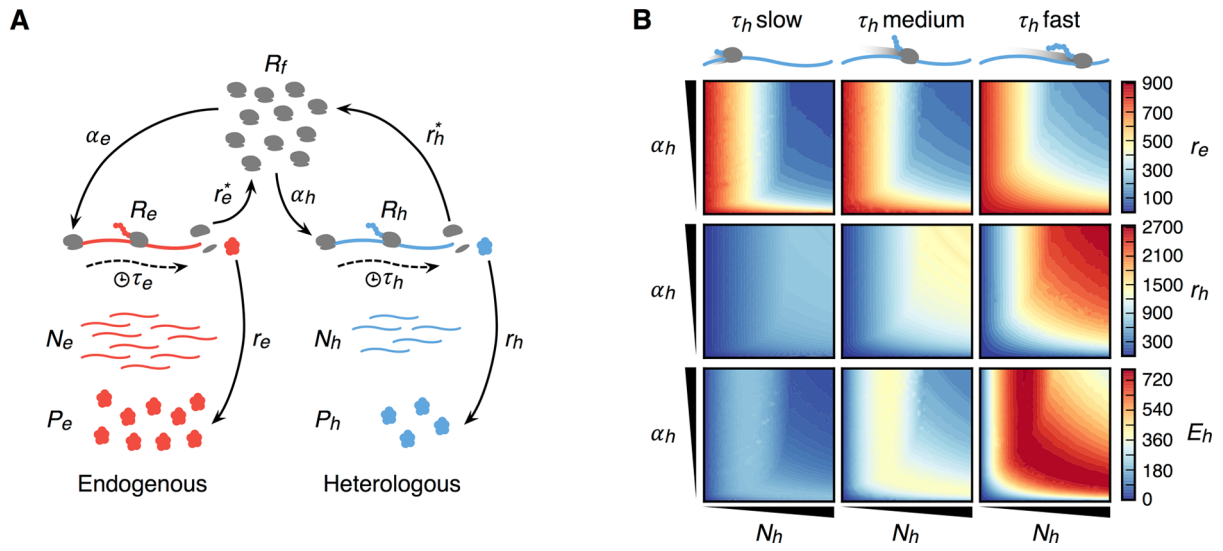


Figure 1. Model of ribosome allocation for cells simultaneously expressing endogenous and heterologous genes. (A) Model overview with key rates and concentrations. R_f, R_e, and R_h denote the ribosome concentrations (ribosomes per cell) that are free, engaged in translation of endogenous mRNAs, and engaged in translation of heterologous mRNAs, respectively. α_e and α_h are the translation initiation rates per second, r_e^* and r_h^* the ribosome recycling rates back to the free pool (ribosomes per second), r_e and r_h the protein production rates (proteins per second), τ_e and τ_h the times in seconds to translate an mRNA, N_e and N_h the mRNA concentrations (mRNAs per cell), P_e and P_h the protein concentrations (proteins per cell) produced for the endogenous and heterologous genes, respectively. (B) Simulation results from the model for varying translation initiation rates (α_h) from 0 to 0.002 s^{-1} and heterologous mRNA levels (N_h) from 5 to 1000 transcripts. Each column corresponds to a heterologous gene that differs in its translation elongation time (τ_h) (slow, 34 s; medium, 17 s; fast, 8.5 s). Rows of the heat maps correspond to the rate of endogenous protein production (r_e) (top row), the rate of heterologous protein production (r_h) (middle row), and the effective rate of heterologous protein production (E_h) in proteins per second units (bottom row).

often stochastic nature of these models make efficient simulation a challenge. Furthermore, they rely on several parameters, such as decoding rates for individual codons, which are often not known and can potentially change due to codon-context effects.[15] This hampers the utilization of these models for detailed explorations of potential design spaces and makes quantitative predictions a challenge.

In contrast, minimal "toy models" of physical systems have been extensively used to gain a deeper understanding of the emergence of complex behaviors in physical systems.[16] One of the most well-known is the Ising model developed to understand the potential phase transitions of magnetized media.[17] Such models purposefully ignore some details of the system, distilling only features of the system that are critical for its behavior. Such idealized models have been built to study the connections among nutrient availability, gene expression, and growth of bacterial cells.[18,19] These have shown striking conserved links between these factors and inherent trade-offs that cells face because of their finite internal resources. However, these models are still of significant complexity and contain large numbers of unknown parameters that must be fitted. This makes them slow to simulate and hampers their use in design-based tasks of synthetic constructs where large design spaces may need to be explored.

To address this limitation, we took a similar minimalistic approach and modeled the dynamic allocation of ribosomes in cells under combined varying protein synthesis demands and environmental stress. Unlike existing models comprising a myriad of equations and parameters, our minimal model is described by only two dynamical equations with eight parameters in total. It exploits delay differential equations (DDEs) to capture the critical time during which a ribosome is engaged in translation of a gene and so is not available to the rest of the cell. Despite its mathematical simplicity, the model faithfully reproduces experimentally measured shifts in protein production rate for a variety of combined expression burdens and temperature stresses. Furthermore, because each model parameter corresponds directly to a key design feature of the synthetic expression construct, it allows prediction of the impact of various design features on shared protein synthesis resources and cell viability. This makes the model useful for performing detailed *in silico* studies of design spaces to find synthetic constructs that maximize productivity while minimizing their impact on endogenous cellular processes.

RESULTS

A Minimal Model of Ribosome Allocation Dynamics. In our model, cells contain a fixed total concentration of ribosomes (R_t) that are split into the following fractions: (i) free and not engaged in translation (R_f) and (ii) translating endogenous (R_e) or heterologous (R_h) mRNAs (Figure 1A). By conservation of ribosome concentrations, $R_f = R_t - (R_e + R_h)$. An "averaged" endogenous transcript was used to capture the native protein synthesis demands. Because mRNA turnover is generally much faster than that of a protein, the mRNA concentrations of endogenous (N_e) and heterologous synthetic genes (N_h) were assumed to remain constant at steady state levels (Table 1). Dynamics of the concentrations of the translating ribosomes for the endogenous and heterologous genes were described by two DDEs:

$$\frac{dR_e}{dt} = R_f(t)\alpha_e N_e - R_f(t-\tau_e)\alpha_e N_e \quad (1)$$

$$\frac{dR_h}{dt} = R_f(t)\beta_h(t)N_h - R_f(t-\tau_h)\beta_h(t-\tau_h)N_h \quad (2)$$

where α_e is the translation initiation rate for endogenous genes (Table 1) and $\beta_h(t)$ is the effective translation initiation rate for

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Table 1. Model Parameter Values

name	description	value	units
R_t	total ribosome concentration	26300[a]	ribosomes per cell
N_e	endogenous mRNA concentration	4140[b]	mRNAs per cell
N_h	heterologous mRNA concentration	5−1000	mRNAs per cell
α_e	translation initiation rate of endogenous mRNA	1.9×10^{-5}[c]	inverse seconds
α_h	translation initiation rate of heterologous mRNA	1.9×10^{-7} to 1.9×10^{-3}	inverse seconds
τ_e	translation time of an endogenous mRNA	17[d]	seconds
τ_h	translation time of a heterologous mRNA	8.5, 17, 34	seconds
s_h	maximal no. of ribosomes per heterologous mRNA	38[e]	ribosomes

[a]Taken from ref 50. [b]Each gene is considered as a separate mRNA, and we assume 1380 mRNAs present in *Escherichia coli* cells under normal growth conditions[50] with each containing three genes.[51] [c]Calculated assuming 0.5 ribosome initiation per second for endogenous genes and a ribosome concentration of 26300 ribosomes per cell.[50] [d]Calculated assuming a rate of 20 amino acids per second for translation[50] and an average length of 340 amino acids per gene. [e]Assumes a ribosome covers ∼9 codons.

heterologous genes that can vary over time. We assume that once a ribosome initiates, translation of the protein takes a constant time given by τ_e and τ_h for endogenous and heterologous genes, respectively (Table 1). Equations 1 and 2 capture the difference in the rates at which the free ribosomes initiate and elongating ribosomes terminate, after the appropriate constant translation time delay.

Because ribosome collisions are rare for endogenous genes,[20] we assume that translation initiation will always be limiting, and thus, α_e is assumed to remain constant. However, as heterologous constructs can implement very fast initiation rates or long translation times, there is the potential for ribosomes to collide and queue. To account for such events, eq 2 includes a variable effective translation initiation rate defined as

$$\beta_h(t) = \min\left\{\alpha_h, \left[1 - \frac{R_h(t)}{N_h s_h}\right]\frac{s_h}{\tau_h}\right\} \quad (3)$$

where s_h is the maximal number of ribosomes that can simultaneously translate the mRNA. Smaller s_h values correspond to mRNAs that either contain early ribosome queuing points or have elongation dynamics that restrict the maximal achievable ribosome density. As heterologous genes are often codon-optimized for efficient translation elongation in the host of interest, they will not contain an excessive clustering of slow codons. We therefore assume s_h is close to the maximal possible density, which can be easily calculated given the length of the gene and the ∼9-codon footprint of a ribosome (Table 1). This assumption can be relaxed if further information about the translation elongation dynamics of the gene is known, allowing s_h to be set to a gene specific value.

The effective initiation rate will in most cases be limited by initiation rate α_h of the ribosome binding site, giving $\beta_h(t) = \alpha_h$. However, if faster initiation rates are considered, there is a chance that the translation elongation rate s_h/τ_h becomes limiting.[21,22] Under this regime, the ribosome density along the heterologous mRNAs will grow, increasing the chance of a collision between free ribosomes attempting to initiate and those translating the mRNA. The second term in the "min" function of eq 3 captures the potential for this to happen. If no ribosomes are translating the heterologous mRNAs, then free ribosomes can initiate at a maximal rate (i.e., the rate at which a ribosome site is translated), giving $\beta_h(t) = s_h/\tau_h$. When the density of translating ribosomes increases, so too does the chance of a collision during initiation. To account for this, we assume that the chance of a successful initiation is proportional to the availability of the initiation site. Therefore, as translating ribosomes can potentially occlude this site, we modify the effective initiation rate by the factor $1 - [R_h(t)/N_h s_h]$, making it inversely proportional to the ribosome density.

We performed approximately 22500 simulations to predict the behavior of protein synthesis in cells exposed to a wide range of biologically realistic expression demands of a synthetic protein. We systematically varied the number of mRNAs (N_h), the translation time (τ_h), and the translation initiation rate (α_h) of the heterologous gene, while maintaining endogenous protein production parameters constant at biologically realistic values from the literature (Table 1). These three variables of the heterologous gene expression were chosen because they directly relate to common parameters that can be modulated when designing synthetic expression constructs, namely, mRNA expression level, strength of translation initiation, and elongation dynamics through modulation of the codon usage choice. From these simulations, the steady state rates of endogenous and heterologous protein production, r_e and r_h, respectively, were inferred from the rates at which ribosomes completed translation of each type of mRNA. Because the actual rate of heterologous production is closely tied to the ability of the host to maintain endogenous protein levels, we defined an effective heterologous production rate as

$$E_h = r_h\left(\frac{r_e}{r_e^-}\right) \quad (4)$$

where r_e^- is the endogenous production rate when no heterologous construct is present. It has been shown that suppression of endogenous gene translation (e.g., through production of a heterologous protein) has a negative effect on growth that takes the form of a linear trade-off between these two factors.[12] Although we do not capture cellular growth in our model, by modulating the effective production rate by the fractional drop in normal endogenous translation r_e/r_e^-, we ensure that E_h provides a comparable measure of relative heterologous protein production even when growth is impacted. Furthermore, if the model is fitted to data in which growth rate is affected, then an additional constant of proportionality can be added to make E_h a measure of the absolute protein production rate.

Simulations displayed two key behaviors. First, endogenous and heterologous expression rates exhibited opposing effects because of competition for the shared ribosome pool (Figure 1B, top and middle rows). Second, the effective protein production rate (E_h) displayed a nonlinear response to the initiation rate and the number of heterologous mRNAs (Figure

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

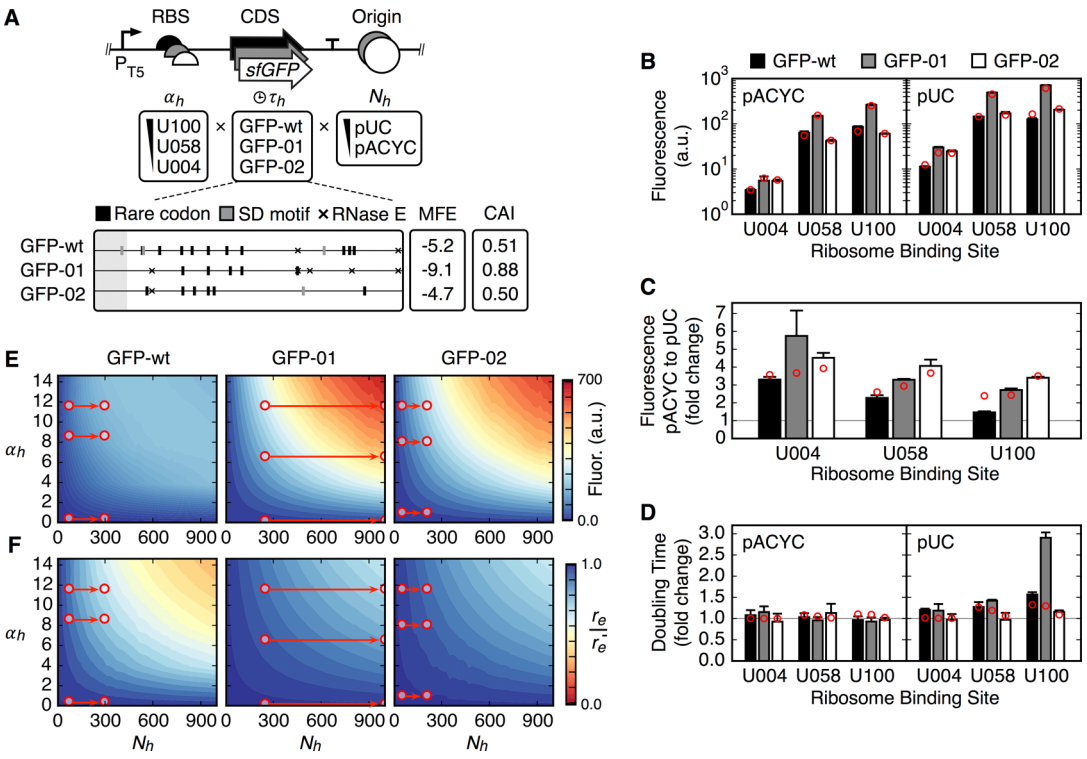


Figure 2. Synonymous sfGFP codon variants with varying transcription and translation levels differently impact expression and cell growth. (A) Design of sfGFP expression plasmids with transcript numbers N_h, translation initiation rate α_h, and translation elongation time τ_h varied by altering the origin of replication, RBS, and CDS, respectively. The close-up shows a comparison of the three sfGFP variants. Light gray regions cover the 5′ UTR; black bars depict the positions of rare codons, gray bars SD-like sites, and black crosses RNase E cleavage sites within the gene. MFE represents the mean folding energy for the −4 to +37 bp region around the start codon in kilocalories per mole, and CAI is the codon adaptation index. (B) sfGFP variant expression 1.75 h after induction (see Figures S1 and S2 for time series data). sfGFP fluorescence was normalized to the biomass. Red circles denote the values fitted by the model. (C) Fold change in fluorescence for each sfGFP variant and RBS between low-copy number (pACYC) and high-copy number (pUC) plasmids with the model fit shown by red circles. (D) Doubling time of each variant normalized to cells without a plasmid. Data are means ± SD of four biological replicates. Red circles show the predicted fold change decrease in endogenous expression rate when the synthetic construct is present. (E) Predicted protein expression (i.e., arbitrary fluorescence, Fluor.) heat maps generated from the model. (F) Fraction of endogenous expression rate achieved compared to when no synthetic gene is present (i.e., r_e/r_e^-). Red circles denote the location of each expression construct. Circles are grouped into pairs corresponding to the RBS (top, U100; middle, U058; bottom, U004). The arrows show the movement from the low-copy number (pACYC) to high-copy number (pUC) plasmid. α_h is given in units of 10^{-5} s^{-1}.

1B, bottom row). This arose from a trade-off in diverting resources to the foreign synthetic product and the need to ensure basal levels of endogenous protein production to maintain basic cell functions.

Systematic Experimental Exploration of Expression Space. To experimentally verify the theoretically identified responses, various combinations of synthetic genetic parts were used to construct expression vectors with precisely controlled transcription and translation of a heterologous superfolder GFP (sfGFP) protein[23] (Figure 2A). This protein was chosen because of its robust and rapid folding that allows for fluorescence measurements to act as an accurate measurement for protein expression.[24] Transcript levels (N_h) were modulated using a T5-inducible promoter in both low-copy number (pACYC) and high-copy number (pUC) plasmids. The translation initiation rate (α_h) was varied by using three ribosome binding sites (RBSs) with relative strengths of 4% (U004), 58% (U058), and 100% (U100). To minimize contextual effects that can arise between the specific combination of promoters, RBSs, and coding sequences (CDSs), bicistronic RBSs designed by the BIOFAB were employed.[25] These bicistronic parts contain a strong upstream RBS that channels ribosomes to a second downstream RBS used to initiate translation of the CDS. The increased ribosome flux generated in this tandem design helps reduce the extent of formation of secondary structures at the RBS−CDS part junction, improving the consistency of translation initiation across genetic contexts.[26] Finally, the translation time of the sfGFP protein (τ_h) was modified through application of two algorithms optimizing the gene sequence by alternative choices of synonymous codons for the *sfGFP* gene. Such differences can strongly influence translational dynamics and the overall translation elongation time,[27] while maintaining an identical amino acid sequence of the protein product.

A full combinatorial library of 18 sfGFP expression variants (Figure 2A) was constructed, and the expression in *Escherichia coli* cells was characterized using a BioLector microbioreactor platform. The BioLector provides a precisely controlled environment that allows high oxygen transfer and rapid sampling of both biomass and sfGFP fluorescence.[28] In total, we recorded approximately 40000 data points with little variation between biological replicates (Figures S1 and S2). From the exponential growth phase, protein expression levels (sfGFP fluorescence normalized by biomass) and cell doubling times were extracted (Figure 2B,D). We fitted our model to the measured expression data and relative initiation rates to assess how well it could account for the observed differences (Table S1). Increases in the level of expression of >2 orders of

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

magnitude were observed between low- to high-copy number plasmids and weak to strong RBSs (Figure 2B). Furthermore, we observed deviations of up to 5.6-fold between sfGFP variants with identical combinations of promoter, RBS, and plasmid copy number.

Interestingly, the sfGFP variants responded differently to increasing strengths of transcription and translation. When comparing GFP-wt to GFP-02 for the strong U100 RBSs, we found that for low-copy number plasmids GFP-wt displayed a significant 1.4-fold higher level of expression ($P = 4.9 \times 10^{-5}$; Welch's *t*-test), while for high-copy number plasmids, this relationship was reversed with a significant 1.6-fold higher level of expression for GFP-02 [$P = 5.0 \times 10^{-5}$; Welch's *t*-test (Figure 2B)]. This shift was also present in the model for the strong U100 RBS with GFP-wt having a level of expression 1.1-fold higher than that of GFP-02 for the low-copy number plasmid, while for the high-copy number plasmid, GFP-02 showed a level of expression 1.3-fold higher than that of GFP-wt. In the model, this arose from a combination of factors. For low-copy number plasmids, even though GFP-wt was translated less efficiently (longer translation time delay), its higher predicted mRNA level led to a level of expression greater than that of GFP-02 (Table S1). However, for high-copy number plasmids where the number of transcripts increased greatly, the less efficient translation of GFP-wt incurred a significant burden on the host. This resulted in a larger drop in the endogenous production rate for GFP-wt compared to that of GFP-02 (Figure 2F), and greater effective expression of the GFP-02 variant under these conditions (Figure 2E). Experiments showed that GFP-01 also saw greater differences in expression between U058 and U100 RBSs compared to both GPF-wt and GFP-02 for stronger RBSs. The model predicts that this is because the expression is limited by low translation initiation rates for GFP-01 (many mRNAs, fast translation), whereas for GFP-wt (few mRNAs, slow translation) and GFP-02 (few mRNAs, fast translation), it is limited by elongation rate.

To validate these predictions, each coding sequence was analyzed (Figure 2A) with a particular focus on features known to influence gene expression, specifically, rare codon clusters that alter translation elongation dynamics,[27,29,30] internal Shine-Dalgarno (SD)-like sequences,[31] RNase E cleavage sites that would decrease mRNA levels,[32] and folding propensity of mRNA in the vicinity of the translation initiation site.[22,33,34] The GFP-wt variant contained the largest number of rare codons and SD-like sequences; in total, 12 sites were found throughout the gene that clustered near the start and end of the coding region as compared to only five and six sites for the GFP-01 and GFP-02 variants, respectively. The enrichment of putative sites associated with ribosomal pausing would increase the translation time (τ_h) of this variant and consequently lead to a lower level of expression (Figure 1B, middle row), as corroborated experimentally (Figure 2B).

Our model accounts for changes in the effective translation initiation rate (E_h) by considering the density of ribosomes along an mRNA and probability of a collision occurring during initiation. Hence, we calculated the relative initiation rate of each RBS to the strongest U100 part across all sfGFP variants. Notably, previously characterized initiation rates were maintained for low-copy number plasmids,[25] but at high copy numbers, GFP-wt saw a significant increase in the initiation rates of both U004 and U058 that was likely due to a drop in the effective initiation rate of U100 (Figure S3). This resulted in U058 having a relative initiation rate that was ∼18% greater than that of U100. Such a change is consistent with the GFP-wt variant being inefficiently translated such that ribosomes are sequestered for long periods of time (Table S1). This would have a significant impact on endogenous expression for the high-copy number plasmids (Figure 2F) and lead to reduced relative initiation rates for stronger RBSs. Although a similar effect would be present for GFP-01, the efficient translation of this variant ensures rapid recycling of ribosomes. This leads to larger numbers of free ribosomes and so less impact on initiation rates of both endogenous and heterologous genes (Figure 2F).

To improve our understanding of the influence of the transcript levels on heterologous protein production, for each construct we compared its expression between low- and high-copy number variants. The model predicts that synthetic gene variants can potentially respond to increasing transcript levels in different ways. First, if the gene is inefficiently translated (i.e., long translation time), then high mRNA levels cause a large drop in free ribosome numbers as they are unavailable to the host for a long period of time. This severely impacts endogenous expression, slows cell growth, and also reduces the effective rate of expression for the synthetic gene. Second, if the gene is efficiently translated (i.e., short translation time), then high mRNAs levels will have a lesser affect on the host.

In all cases, the experimentally measured fold change in expression from low- to high-copy number variants decreased with increasing RBS strength (Figure 2C). Interestingly, the degree of this change was nonuniform across variants. GFP-wt and GFP-01 showed large fold change decreases, while GFP-02 exhibited marginal changes in fold change expression. This is explained by the model; with an increasing RBS strength, the switch from low to high mRNA levels causes a greater drop in free ribosome numbers. This impacts endogenous expression, slows cell growth, and therefore reduces the effective rate of expression for the synthetic gene. The model predicts that the more marginal changes seen with GFP-02 are due to a combination of low mRNA and fast translation leading to a reduced level of ribosome sequestration.

Fitted expression levels closely matched the experiments with a <0.3-fold difference and a highly significant correlation between the observed and fitted values [$R^2 = 0.997$; $P = 6.2 \times 10^{-19}$ (Figure 2B and Figure S7A)]. The model also captured the increasing fold change difference in expression between GFP-01 and both GFP-wt and GFP-02 variants when moving from the weak U004 to strong U100 RBS (Figure 2B). This was clearly evident for the GFP-01 and GFP-02 variants where for the U004 RBS both had similar expression levels for low- and high-copy number plasmids, whereas for the U100 RBS, the GFP-01 variant had expression levels 4.4- and 3.5-fold higher than that of GFP-02 for the low- and high-copy number plasmids, respectively. These closely matched the 4.1- and 2.9-fold changes, respectively, given by the model (Figure 2B). Furthermore, both experimental and model fits showed that the GFP-02 variant was able to maintain a larger fold change difference in expression between low- and high-copy number plasmids for strong U058 and U100 RBSs (Figure 2C).

In addition to expression of synthetic genes, the model also captures the expression rate of the host's own endogenous genes (Figure 2F). As the ability for a cell to express endogenous genes is closely tied to its growth rate,[12] we compared the rate of endogenous gene expression when no synthetic construct was present to the changes in growth rate

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

for expression constructs where sfGFP variants were strongly expressed [i.e., for the U100 RBS and pUC origin (Figure 2D)]. From the model, we found that GFP-wt and GFP-01 displayed large drops of 33 and 30%, respectively, whereas GFP-02 had an only 9% lower expression rate for endogenous genes (Figure 2D). GFP-02 had the smallest impact on growth rate across all expression variants (Figure 2D). These broad differences were similar to the experimentally measured increases in cell doubling times (Figure 2D). Moreover, the rapid increases in doubling times for the GFP-wt and GFP-01 variants as RBS strength increased for the high-copy number plasmids (Figure 2D) were also observed in the model (Figure 2F).

Although the model is fitted purely to the expression data, the direct link between parameters in the model and properties of the synthetic expression constructs allowed us to indirectly infer several additional attributes. We estimated translation time delays (τ_h^*) of 34, 8.5, and 9.4 s and numbers of transcripts for the low-copy number variants (N_h^*) of 73, 248, and 52 for GFP-wt, GFP-01, and GFP-02, respectively. Furthermore, an increase by a factor of 4.0 for all constructs was predicted for transcript levels between expression from low- and high-copy number plasmids. This ordering of the translation time delays correlates with the number of detected pausing sites (Figure 2A), and the high GFP-01 transcript levels would account for its observed increased level of expression (Figure 2B). To validate these predicted differences in mRNA levels, quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) measurements were taken. These showed that GFP-01 had mRNA levels 3.3- and 4.4-fold higher than those of GFP-wt and GFP-02, respectively, closely matching the 3.4- and 4.7-fold differences, respectively, predicted by the model (Table S1).

Using these fitted translation times, we were also able to generate global expression landscapes (Figure 2E). It was clear to see that translation time delay τ_h strongly determines the ability for other design parameters to modulate expression. For example, as τ_h increases moving from GFP-01 to GFP-wt, faster initiation rates (α_h) have a weaker effect, reducing the feasibility of this parameter to be used to tune expression. It also provides direct insight into how quantitative changes in the major parameters of the expression construct will affect overall production, making this model a versatile tool for optimization of expression and ensuring a minimal burden on the host chassis.

Impact of Additional Stress on Expression Dynamics. Next, we explored the impact of additional abiotic stress (e.g., temperature stress) to assess the suitability of the model for capturing the allocation of translation resources for cells experiencing combined stress. Temperature was chosen as it is often varied to improve the production of properly folded proteins in heterologous hosts.[35] It also poses a problem in large-scale fermentations where unavoidable heterogeneity during mixing within a bioreactor leads to regions with nonoptimal temperatures.[4]

We subjected cells growing first at 37 °C to a range of temperature variations, from 7 to 57 °C (Figure 3A). Different expression times at each temperature were chosen to achieve comparable total expression amounts. Because each of the processes of sfGFP biogenesis (e.g., translation, folding, and aggregation) are temperature-dependent, the expression time was increased at low temperatures and reduced at high temperatures. At the same time, cells were subjected to a large synthetic protein production burden by expressing the

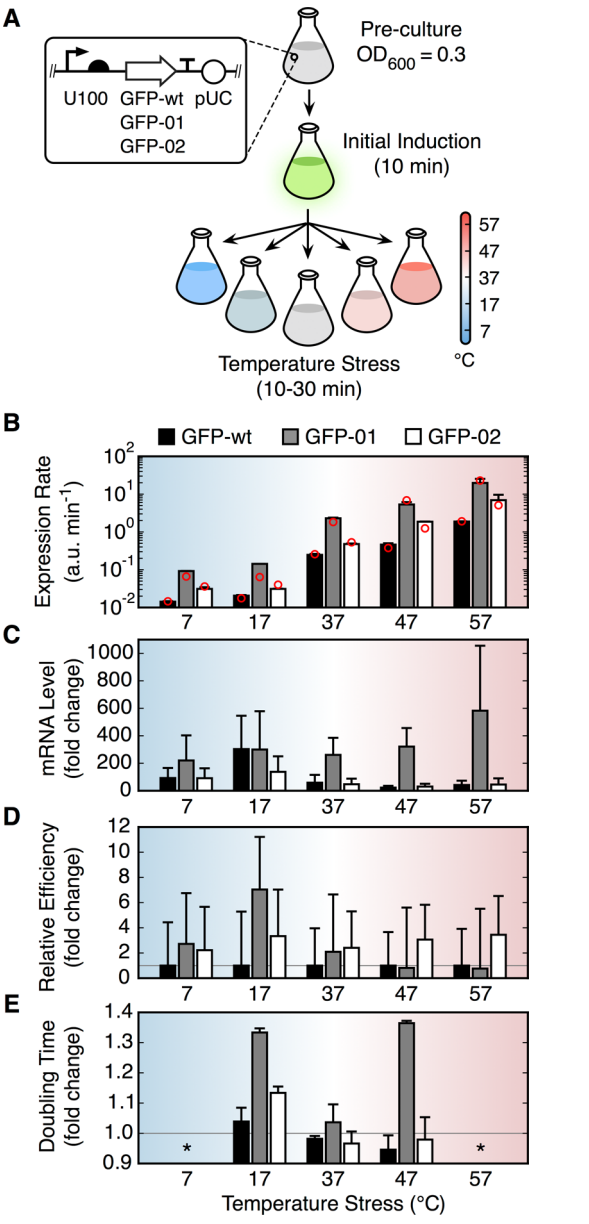


Figure 3. Impact of heat stress on protein expression and growth for cells expressing the sfGFP variants. (A) Overview of the expression construct and experimental setup. (B) Expression of the three sfGFP variants at different temperatures calculated as a change in the fluorescence during temperature stress for the time of stress exposure (7 and 17 °C for 30 min, 37 and 47 °C for 20 min, and 57 °C for 10 min). Red circles denote the model fit. (C) Relative mRNA levels of the sfGFP variants normalized to the expression levels of the endogenous GAPDH mRNA. (D) Relative expression efficiency (fluorescence/mRNA) for each sfGFP variant normalized to that of GFP-wt for each temperature separately. (E) Doubling time of cells expressing different variants at each temperature normalized by cells without a plasmid. Data are means ± SD of two biological replicates. An asterisk indicates no growth was detected for cells exposed to 7 and 57 °C (Figure S5).

sfGFP variants on high-copy number plasmids under the strong U100 RBS (Figure 3B,E). Because temperature has wide-ranging effects on protein synthesis rates and mRNA turnover,[36] we quantified mRNA expression by qRT-PCR and measured the total and folded sfGFP production by immunoblot and FACS, respectively (Figures S4 and S6).

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

Importantly, the fraction of the fluorescent folded sfGFP from the total amount of synthesized sfGFP remained nearly unchanged (Figure S4), suggesting that exposure to different temperatures did not change the folding ability or stability of the sfGFP variants. This means fluorescence faithfully represents protein production after exposure to a temperature down- or upshift.

We next compared the relative mRNA levels for each sfGFP variant. Transcript levels exhibited some variation with temperature (Figure 3C), although a similar ordering in levels between the sfGFP variants was seen throughout: at lower temperatures, these differences were reduced. Conversely, at temperatures of ≥37 °C, GFP-wt and GFP-02 maintained similar levels, while much higher levels that increased with temperature were observed for GFP-01. At 57 °C, the differences in mRNA levels between GFP-wt/GFP-02 and GFP-01 were found to be statistically significant ($P = 0.014$ and 0.037, respectively; Welch's t-test). As changes in expression were fairly uniform across all sfGFP variants, variability in mRNA levels between variants implies differences in relative translation rates per mRNA (Figure 3D). While both GFP-wt and GFP-01 displayed similar efficiencies, the larger increase in the level of expression for the GFP-02 variant at relatively constant mRNA implies a >3-fold higher production rate per mRNA. For the strongly expressing GFP-01 variant, this observation is consistent with expression reaching the cellular capacity boundaries, which is further supported by the large increase in the doubling time of cells at 47 °C (Figure 3E). At 7 °C, we found similar relative efficiencies for all variants when compared to that at 37 °C (Figure 3D), although much lower expression rates (Figure 3B). As growth at this temperature stops (Figure 3E), these similarities are likely due to the cells maintaining a snapshot of their state with minimal new protein synthesis. At 17 °C, we observed a large increase in the relative efficiency of the GFP-01 variant due to a much higher mRNA level for the GFP-wt.

Finally, we fitted our model to these new data by using the previously determined translation time delays (τ_h^*) and separately fitting expression conversion factors (F_x) that account for changes in protein synthesis capacity between the different temperatures (Methods). An excellent fit was obtained with a <0.6-fold difference and a highly significant correlation between the observed and fitted values [$R^2 = 0.992$; $P = 3.5 \times 10^{-13}$ (Figure 3B and Table S2)]. This gave $F_x = 0.21$ (7 °C), 0.17 (17 °C), 5.4 (37 °C), 12.4 (47 °C), and 17.9 (57 °C). These factors closely match known changes in translation rates at different temperatures.[36] For the upshift from 37 to 47 °C, we detected a 2.3-fold change, and for the downshift from 37 to 7 °C, we detected a 26-fold change; both closely match the ∼2- and ∼20-fold changes in ribosome velocities, respectively, for these temperatures as calculated from the literature.[36]

DISCUSSION

We have developed a minimal model of ribosome allocation dynamics that accurately captures the shifts in overall expression under combined protein production demands and ambient stress and embodies experimentally observed trade-offs in expression and growth. The model is able to make precise predictions of transcript levels and translation elongation times from protein expression levels alone. This ability to infer multiple unknown parameters about such systems from expression levels alone offers a novel approach to determining difficult to observe processes.

We purposefully neglected any explicit representation of growth in our current model to simplify its representation and focus on the role of ribosome allocation during stress conditions. Our aim was to assess the behavior of cells under acute stress (i.e., short exposures to temperature shifts), which has an insignificant effect on growth. For longer-term exposure, the effects on growth should be incorporated as several empirical relationships have shown that the capacity for heterologous gene expression, cellular growth, and cell size are closely connected.[12,37]

The use of time delays in dynamic mathematical models of biological systems has been shown to be capable of inducing many interesting behaviors such as oscillations[38,39] as well as influencing stochastic effects.[40] Nevertheless, their use in modeling biology has remained limited. Our model relies on the use of time delays to capture the time it takes for a ribosome to fully translate an endogenous or heterologous transcript. By using a delay to characterize this aspect, we are able to group together the microscopic steps of translation, which significantly simplifies the mathematical description while still ensuring that appropriate dynamical behaviors are maintained. This removes the need for many parameters, such as translocation rates at single codons,[7,8] which are often not known and can potentially vary.[15] Furthermore, these simplifications permit simulations to run several orders of magnitude faster than more common ribosome flow-based models,[7,41] allowing for entire design spaces to be efficiently simulated and explored.

A number of other models have been developed to investigate competition for shared cellular resources. These reveal trade-offs in expression, queuing, and shifts in dynamic behaviors related not only to transcriptional and translational processes[9,18,42−44] but also to degradation machinery[45] and metabolic enzymes.[46] In terms of protein synthesis, several simplified models of transcription and translation have been derived to study the potential indirect effects between multiple types of synthetic genes. This has revealed subtle differences in the ways that trade-offs are manifested when varying the strength of transcription and translation[42,44] and shown that stochastic fluctuations in ribosome numbers can lead to strong anticorrelations between different transcript types as this resource becomes overloaded.[43] Additionally, models capturing translation elongation dynamics have highlighted the role of bottlenecks arising from codon choice that lead to ribosome "traffic jams". These limit gene expression and place an increased burden on the host due to sequestering of this shared resource.[7−9,20,41]

Most of these previous models focus exclusively on expression of synthetic constructs. Unlike our minimal model, they do not directly capture the potential impact on endogenous expression and consequently the effects on growth. Furthermore, many models assume that translation occurs at a specific rate, ignoring the fact that for strongly expressed synthetic genes the large number of mRNAs will lead to a sizable fraction of the ribosome pool being physically sequestered. The model of Ceroni et al.[9] does include more detailed consideration of translational dynamics, capturing a free ribosome pool and translational dynamics along an mRNA. However, this comes at the cost of increasing complexity and the need to provide translation rates of every codon. This is a real challenge and can vary for differences in local codon composition.[15] In contrast, our simple model requires information regarding the translation time of a gene, which

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

can be inferred with relative ease from a small expression library due to the small number of free parameters. This has the additional benefit of providing single parameters to capture the key role of transcription, translation initiation, and translation elongation dynamics and thus clear links to design parameters of an expression construct.

Our focus in this work has been the experimental validation of our minimal model to observed shifts in expression and growth. Having reproduced these behaviors, the model can be extended in a several new ways for applications in bioengineering. The simplicity of our model allows relatively small libraries of expression constructs to provide sufficient information to infer unknown parameters relating to mRNA concentrations, translation initiation rates, and protein translation times, which are often a challenge to accurately measure. Moreover, once the protein translation time is known, it is possible to quickly generate the entire predicted expression space and find specific quantitative values for each parameter that will result in a desired expression level and impact on the host. This is of particular interest when designing genetic circuits or metabolic pathways where the precise tuning of expression is essential for correct function,[47] but where it is also preferable to avoid an unnecessary burden on the host.[9] Furthermore, these applications are supported by the ability to efficiently simulate these models, permitting large design spaces to be quickly evaluated.

The introduction of synthetic circuits and pathways for the production of protein products and metabolites remains a vital goal of industrial applications, and their efficiency becomes increasingly important as the scale of these systems grows. The burden a genetic circuit places upon the host can be alleviated through better use of shared cellular protein synthesis resources,[9] yet understanding the precise changes in optimizing the DNA design remains challenging. Our model connects experimentally tunable design parameters with quantitative predictions that are directly actionable. This makes the model a valuable resource for efficiently exploring the optimal trade-offs between these properties and allows for targeted changes to constructs that improve synthetic gene expression, while simultaneously considering additional demands that may arise from environmental factors and limited cellular resources.

METHODS

Expression Vectors. Two plasmids formed the basis for our expression vectors: pJ421 (Kan^R; DNA2.0, Newark, CA), which contains a pACYC low-copy number origin of replication, and pJ441 (Kan^R; DNA2.0), which contains a pUC high-copy number origin of replication. Both vectors bear an isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible T5 promoter. Three RBS sequences developed and characterized by the BIOFAB[25] were chosen and inserted upstream of the gene being expressed. The expression vectors and RBS elements were combined to generate a full combinatorial library driving expression of three sfGFP synonymous codon variants. In addition to a wild-type GFP-wt sequence,[24] two codon optimization algorithms were applied to design optimized sfGFP variants that retained the same amino acid sequence. Constraints were imposed during this process to ensure no interference with BsaI restriction sites required for cloning. The first optimization method used a codon pair optimization approach described in U.S. Patent 8,812,247 B2 (Method for achieving improved polypeptide expression) to produce the GFP-01 variant, and a second variant GFP-02 was produced using an approach based on a proprietary algorithm developed by DNA2.0 that applies a machine-learning technique on internally collected expression data sets. Sequences for all sfGFP variants and RBSs are provided in Tables S3 and S4. Expression elements and sfGFP genes were synthesized by DNA2.0. Cloning was performed using *E. coli* 10-β strains (New England Biolabs).

BioLector Experiments. BioLector experiments were conducted using *E. coli* RV308 (ATCC #31608; Su-, lac X 74, gal ISII: OP308, strA) cells transformed by the combinatorial expression library and grown in Luria-Bertani medium supplemented with casamino acids [LBC; 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, and 1 g/L casamino acids (pH 7.0)]. Cells were cultured in the BioLector microbioreactor platform (m2p-laboratories GmbH, Aachen, Germany). Biomass concentrations were measured via scattered light at 620 nm and sfGFP fluorescence at 520 nm (485 nm excitation). Common gains of 20 and 40 were used for the biomass and sfGFP measurements, respectively. Starter cultures were grown from single colonies in LBC medium supplemented with 50 μg/mL kanamycin at 37 °C overnight. These were then diluted 100-fold in LBC medium containing the same concentration of antibiotic, and expression was performed in 48-well FlowerPlate microtiter plates (m2p-laboratories GmbH, part number MTP-48-B). We used 1 mL culture volumes shaken at 900 rpm. Humidity control was permitted on the BioLector, and biomass and sfGFP readings were taken every 6 min. All experiments were performed with four biological replicates.

Temperature Stress Experiments. GFP-wt, GFP-01, and GFP-02 genes with the same U100 RBS were cloned in pJ441 plasmids with a pUC origin and expressed in *E. coli* MG1655 (ATCC #700926) cells. Starter cultures from single colonies were grown overnight in Luria-Bertani (LB) medium supplemented with 50 μg/mL kanamycin at 37 °C and 200 rpm. These cultures were diluted 100-fold and used to inoculate fresh LB medium. Expression of the sfGFP constructs was induced once an OD_{600} of 0.3 was reached using 1 mM IPTG for 10 min. The cells were harvested at 4000 rpm for 4 min at room temperature, and the pellet was dissolved in preheated or precooled fresh LB medium and further incubated at different temperatures (7 and 17 °C for 30 min, 37 and 47 °C for 20 min, and 57 °C for 10 min). Expression of sfGFP was measured by flow cytometry on a FACSCalibur-Sort instrument (BD Biosciences) equipped with a 15 mW argon ion laser (488 nm) for excitation. The instrument settings were in logarithmic mode: FSC-H, E01; SSC-H, 375; FL1-H, 674; FL2-H, 706; FL3, 716. To eliminate nonbacterial particles, a threshold value of 217 was set on SSC. The cell culture was diluted in 0.22 μm filtered PBS buffer [58 mM Na_2HPO_4, 17 mM NaH_2PO_4, and 68 mM NaCl (pH 7.4)] to a density of approximately 10^6 cells/mL. For each culture, 50000 events were collected. The total cell population in a sample was gated in a bivariate dot plot of FSC and SSC. FlowJo (version 10.0.7r2) software was used for analysis. Nontransformed cells were used to normalize the autofluorescence for each condition.

In Gel Fluorescence and Immunodetection. Cell aliquots for which OD_{600} = 3 were treated as described by Geertsma et al.[48] The samples were shaken using a Retsch MM 400 device for 30 at 25 s^{-1}. After being mixed with 5× sodium dodecyl sulfate (SDS) loading buffer [120 mM Tris-HCl (pH 6.8), 2% (w/v) SDS, 50% (v/v) glycerol, 100 mM DTT, and 0.1% (w/v) bromophenol blue], the protein samples were

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

resolved on a 10% SDS−polyacrylamide gel. In gel sfGFP fluorescence was visualized immediately using the Bio-Rad ChemiDoc MP imaging system. Using the same gels, sfGFP variants were detected by immunoblotting using anti-GFP (1:2000; Roche) as a primary antibody and visualized by chemiluminescence via a secondary goat anti-mouse antibody coupled with horseradish peroxidase (1:10000; Bio-Rad). Detection was performed using the Bio-Rad ChemiDoc MP imaging system. For total protein sample analysis, cell aliquots were treated as described by Hess et al.[49] The samples were shaken with a Retsch MM 400 device as described above instead of the sonication step. The protein samples were mixed with 5× SDS loading buffer [313 mM Tris-HCl (pH 6.8), 5% (w/v) SDS, 50% (v/v) glycerol, 100 mM DTT, and 0.5% (w/v) bromophenol blue] and heated at 95 °C for 3 min. The samples were resolved on a 10% SDS−polyacrylamide gel. The gels were stained with colloidal Coomassie staining solution [0.08% (w/v) Coomassie Brilliant Blue G250, 8% (w/v) ammonium sulfate, 1.6% (v/v) *o*-phosphoric acid, and 20% (v/v) methanol].

qRT-PCR. Total RNA was isolated using the TRI reagent (Sigma-Aldrich) according to the manufacturer's instructions with an additional washing step using 70% ethanol. The quantity of RNA was determined by the absorbance ratio A_{260}/A_{280} (≥1.8) and denaturing agarose gel electrophoresis. DNA depletion was performed using DNase I (Thermo Scientific), followed by cDNA synthesis using a RevertAid RT Kit (Thermo Scientific) with random hexamer primers. Real-time PCR was performed using the QuantiFast SYBR Green PCR Kit (Qiagen) on an Agilent Technologies Stratagene Mx3005P instrument. GAPDH mRNA was used for normalization.

Model Fitting. To fit the model, we made use of the fine-grained simulations performed to generate Figure 1B. We used linear interpolation between these simulated points to provide a function E_h^* that given any τ_h, N_h, and α_h value would generate an interpolated effective heterologous production rate. This rate was defined to incorporate an additional factor F_x to account for the necessary conversion from expression units in the model (proteins per second) and the experimentally measured fluorescence units (a.u. per second) and to allow for the impact of drops in endogenous production rate to have varying effects on heterologous production. This resulted in

$$E_h^* = F_x r_h\left(\frac{r_e}{r_e^-}\right) \tag{5}$$

where r_e and r_h were interpolated from the simulation data for given τ_h, N_h, and α_h values and r_e^- = 885 proteins/s calculated from the simulation in which no heterologous production was present.

For the BioLector data, we made use of the relationships in the transcription and translation strengths and fitted parameters that allowed these to be converted to relevant units for the model. For low-copy number plasmids, we directly fitted N_h parameters for the mRNA concentrations of each variant. For high-copy number plasmids, we assumed increases in concentration by an identical factor across all variants as only the origin of replication had changed. This resulted in mRNA concentrations for the high-copy number plasmid given by F_pN_h, where F_p is the common factor capturing this increase. For the different RBS strengths, we used the relative strengths calculated from experiment (Figure S3) and fitted a conversion factor F_α to translate this into inverse seconds required for the model. A total of nine parameters were fitted to the 18 fluorescence measurements with the model constrained by the fixed relative strengths of the RBSs and a common factor relating mRNA levels between low- and high-copy number plasmids.

For the temperature stress data, we used fitted τ_h values for each variant from the BioLector data. Transcript concentrations for each variant were given by F_mN_h, where N_h was taken as the relative qRT-PCR datum for each variant and temperature and F_m was a fitting factor shared across all measurements to convert these relative values into the appropriate transcripts/cell units for use in the model. Because the general rate of translation is known to change with temperature,[36] separate factors were used to convert expression rates from the model into experimentally measured expression units. These temperature specific parameters $F_x^{7\ °C}$, $F_x^{17\ °C}$, $F_x^{37\ °C}$, $F_x^{47\ °C}$, and $F_x^{57\ °C}$ were used as a multiplicative factor for the predicted effective heterologous production rate (E_h^*) to convert the value to experimental expression units (a.u. per second). A total of seven parameters were fitted to the 15 fluorescence measurements with the model further constrained by the measured mRNA levels for each variant and temperature.

All data were fitted using a sum of squared errors (SSE) approach and the "minimize" function of SciPy version 0.15.0 that implements a sequential least squares programming (SLSQP) algorithm. Fitted parameter values for the BioLector and temperature stress expression data are provided in Tables S1 and S2.

Model Simulation and Data Analysis. The MATLAB code for the model is available at https://github.com/BiocomputeLab/riboalloc. Model DDEs were solved numerically using MATLAB version 2014b and the "dde23" solver. All data analysis was performed using Python version 2.7.9, NumPy version 1.9.1, SciPy version 0.16.0, and matplotlib version 1.4.2.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.6b00040.

Growth curves for sfGFP variants and expression constructs (Figure S1), expression curves for sfGFP variants and expression constructs (Figure S2), relative RBS strengths (Figure S3), electrophoretic mobility and immunoblotting of sfGFP constructs (Figure S4), growth curves for each sfGFP variant under varying temperature stresses (Figure S5), fluorescence distributions before and after each temperature stress (Figure S6), comparison of measured and fitted expression data (Figure S7), fitted parameters for BioLector data (Table S1), fitted parameters for temperature stress data (Table S2), sfGFP variant sequences (Table S3), ribosome binding site sequences (Table S4), and supporting references (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: thomas.gorochowski@bristol.ac.uk.

Author Contributions

T.E.G., J.A.R., Z.I., I.A.-K., and R.A.L.B. conceived and designed the experiments. T.E.G. developed the mathematical model, performed BioLector experiments, and started initial flow cytometry experiments. I.A.-K. performed temperature stress

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

experiments, flow cytometry, and qRT-PCR. T.E.G., I.A.-K., Z.I., and J.A.R. conducted the data analysis. All authors wrote and edited the manuscript. J.A.R. and Z.I. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

All authors were supported by European Commission-funded Marie-Curie Actions Initial Training Network for Integrated Cellular Homeostasis (NICHE) Project 289384. T.E.G. was additionally supported by BrisSynBio, a BBSRC/EPSRC Synthetic Biology Research Centre (BB/L01386X/1).

■ REFERENCES

(1) Lopez-Maury, L., Marguerat, S., and Bahler, J. (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat. Rev. Genet. 9* (8), 583−93.
(2) Gunasekera, T. S., Csonka, L. N., and Paliy, O. (2008) Genome-wide transcriptional responses of Escherichia coli K-12 to continuous osmotic and heat stresses. *J. Bacteriol. 190* (10), 3712−20.
(3) *Physiological Stress Responses in Bioprocesses* (2004) Vol. *89*, Springer-Verlag, Berlin.
(4) Lara, A. R., Galindo, E., Ramírez, O. T., and Palomares, L. A. (2006) Living with heterogeneities in bioreactors. *Mol. Biotechnol. 34* (3), 355−381.
(5) Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P. C., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H.-Y., Neubauer, P., van der Lans, R., Luyben, K., Vrabel, P., and Manelius, Å. (2001) Physiological responses to mixing in large scale bioreactors. *J. Biotechnol. 85*, 175−185.
(6) Schweder, T., Krüger, E., Xu, B., Jürgen, B., Blomsten, G., Enfors, S.-O., and Hecker, M. (1999) Monitoring of Genes That Respond to Process-Related Stress in Large-Scale Bioprocesses. *Biotechnol. Bioeng. 65* (2), 151−159.
(7) Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E., and Tuller, T. (2011) Genome-scale analysis of translation elongation with a ribosome flow model. *PLoS Comput. Biol. 7* (9), e1002127.
(8) Zia, R. K., Dong, J., and Schmittmann, B. (2011) Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *J. Stat. Phys. 144* (2), 405−428.
(9) Ceroni, F., Algar, R., Stan, G. B., and Ellis, T. (2015) Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods 12* (5), 415−418.
(10) Li, G. W., Burkhardt, D., Gross, C., and Weissman, J. S. (2014) Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell 157* (3), 624−35.
(11) Holcik, M., and Sonenberg, N. (2005) Translational control in stress and apoptosis. *Nat. Rev. Mol. Cell Biol. 6* (4), 318−27.
(12) Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z., and Hwa, T. (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science 330*, 1099−1102.
(13) Dong, H., Nilsson, L., and Kurland, C. G. (1995) Gratuitous Overexpression of Genes in *Escherichia coli* Leads to Growth Inhibition and Ribosome Destruction. *J. Bacteriol. 177* (6), 1497−1504.
(14) Vind, J., Sørensen, M. A., Rasmussen, M. D., and Pedersen, S. (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. *J. Mol. Biol. 231*, 678−688.
(15) Chevance, F. F. V., Le Guyon, S., and Hughes, K. T. (2014) The effects of codon context on *in vivo* translation speed. *PLoS Genet. 10* (6), e1004392.
(16) Marzuoli, A. (2009) Toy models in physics and the reasonable effectiveness of mathematics. *Scientifica Acta 3* (1), 13−24.
(17) Ising, E. (1925) Beitrag zur Theorie des Ferromagnetismus. *Z. Phys. 31* (1), 253−258.
(18) Weisse, A. Y., Oyarzun, D. A., Danos, V., and Swain, P. S. (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U. S. A. 112* (9), E1038−47.
(19) Hui, S., Silverman, J. M., Chen, S. S., Erickson, D. W., Basan, M., Wang, J., Hwa, T., and Williamson, J. R. (2015) Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol. 11*, e784.
(20) Mitarai, N., Sneppen, K., and Pedersen, S. (2008) Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J. Mol. Biol. 382* (1), 236−245.
(21) Plotkin, J. B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet. 12* (1), 32−42.
(22) Kudla, G., Murray, A. W., Tollervey, D., and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in Escherichia coli. *Science 324* (5924), 255−8.
(23) Gorochowski, T. E., van den Berg, E., Kerkman, R., Roubos, J. A., and Bovenberg, R. A. (2014) Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of Escherichia coli. *ACS Synth. Biol. 3* (3), 129−39.
(24) Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol. 24* (1), 79−88.
(25) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods 10* (4), 354−60.
(26) Mutalik, V. K., Guimaraes, J. C., Cambray, G., Mai, Q. A., Christoffersen, M. J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G., Keasling, J. D., Endy, D., and Arkin, A. P. (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat. Methods 10* (4), 347−353.
(27) Kane, J. F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol. 6*, 494−500.
(28) Kensy, F., Zang, E., Faulhammer, C., Tan, R. K., and Buchs, J. (2009) Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. *Microb. Cell Fact. 8*, 31.
(29) Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A., and Roubos, J. A. (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res. 43* (6), 3022−32.
(30) Wohlgemuth, S. E., Gorochowski, T. E., and Roubos, J. A. (2013) Translational sensitivity of the Escherichia coli genome to fluctuating tRNA availability. *Nucleic Acids Res. 41* (17), 8021−33.
(31) Li, G. W., Oh, E., and Weissman, J. S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature 484* (7395), 538−41.
(32) Carpousis, A. J. (2007) The RNA Degradosome of Escherichia coli: An mRNA-Degrading Machine Assembled on RNase E. *Annu. Rev. Microbiol. 61* (1), 71−87.
(33) Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Bluthgen, N. (2013) Efficient translation initiation dictates codon usage at gene start. *Mol. Syst. Biol. 9*, 675.
(34) Goodman, D. B., Church, G. M., and Kosuri, S. (2013) Causes and effects of N-terminal codon bias in bacterial genes. *Science 342*, 475−479.
(35) Weickert, M. J., Doherty, D. H., Best, E. A., and Olins, P. O. (1996) Optimization of heterologous protein production in Escherichia coli. *Curr. Opin. Biotechnol. 7*, 494−499.
(36) Farewell, A., and Neidhardt, F. C. (1998) Effect of temperature on *in vivo* protein synthetic capacity in *Escherichia coli*. *J. Bacteriol. 180* (17), 4704−4710.

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(37) Basan, M., Zhu, M., Dai, X., Warren, M., Sevin, D., Wang, Y. P., and Hwa, T. (2015) Inflating bacterial cells by increased protein synthesis. *Mol. Syst. Biol. 11* (10), 836.
(38) Bratsun, D., Volfson, D., Tsimring, L. S., and Hasty, J. (2005) Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. U. S. A. 102* (41), 14593−14598.
(39) Danino, T., Mondragon-Palomino, O., Tsimring, L., and Hasty, J. (2010) A synchronized quorum of genetic clocks. *Nature 463* (7279), 326−30.
(40) Zavala, E., and Marquez-Lago, T. T. (2014) Delays induce novel stochastic effects in negative feedback gene circuits. *Biophys. J. 106* (2), 467−78.
(41) Zur, H., and Tuller, T. (2012) RFMapp: ribosome flow model application. *Bioinformatics 28* (12), 1663−4.
(42) Gyorgy, A., Jimenez, J. I., Yazbek, J., Huang, H. H., Chung, H., Weiss, R., and Del Vecchio, D. (2015) Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys. J. 109* (3), 639−46.
(43) Mather, W. H., Hasty, J., Tsimring, L. S., and Williams, R. J. (2013) Translational cross talk in gene networks. *Biophys. J. 104* (11), 2564−72.
(44) Carbonell-Ballestero, M., Garcia-Ramallo, E., Montanez, R., Rodriguez-Caso, C., and Macia, J. (2016) Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the Ohm's law. *Nucleic Acids Res. 44* (1), 496−507.
(45) Cookson, N. A., Mather, W. H., Danino, T., Mondragon-Palomino, O., Williams, R. J., Tsimring, L. S., and Hasty, J. (2011) Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol. 7*, 561.
(46) Rondelez, Y. (2012) Competition for catalytic resources alters biological network dynamics. *Phys. Rev. Lett. 108* (1), 018102.
(47) Brophy, J. A. N., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508−520.
(48) Geertsma, E. R., Groeneveld, M., Slotboom, D.-J., and Poolman, B. (2008) Quality control of overexpressed membrane proteins. *Proc. Natl. Acad. Sci. U. S. A. 105* (15), 5722−5727.
(49) Hess, A. K., Saffert, P., Liebeton, K., and Ignatova, Z. (2015) Optimization of translation profiles enhances protein expression and solubility. *PLoS One 10* (5), e0127039.
(50) Bremer, H., and Dennis, P. P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In *Escherichia coli and Salmonella*, Neidhardt, F. C., Ed. ASM Press: Washington DC.
(51) Nunez, P. A., Romero, H., Farber, M. D., and Rocha, E. P. (2013) Natural selection for operons depends on genome size. *Genome Biol. Evol. 5* (11), 2242−54.

DOI: 10.1021/acssynbio.6b00040
ACS Synth. Biol. XXXX, XXX, XXX−XXX