Fitness-driven scaling laws between mRNA and protein levels

Yichen Yan[1] and Jie Lin[1,2,3]

[1] *Peking-Tsinghua Center for Life Sciences, Peking University, Beijing, China*
[2] *Center for Quantitative Biology, Peking University, Beijing, China*
[3] *School of Physics, Peking University, Beijing, China*

(Dated: September 28, 2025)

The connection between protein and mRNA abundance represents a central question in gene expression. While high correlations between mRNA and protein levels are widely observed across different organisms, the evolutionary forces driving these correlations remain elusive. Here, by balancing the cost and benefit of gene expression, we identify the optimal scalings between mRNA and protein levels maximizing cell fitness, where the scaling exponent is positively related to the toxicity effects of protein overexpression. Notably, we predict a lower bound for the exponent, which is 0.5. For the model organism *Saccharomyces cerevisiae*, our model predictions, which incorporate biologically relevant parameters, align well with experimental data. By analyzing genome-wide data across organisms, we find that all organisms exhibit a scaling exponent above the predicted lower bound. These universal phenomena, together with our cost-benefit trade-off model, rationalize the dominance of transcriptional regulation. Our work demonstrates how evolutionary driving forces govern the principles of gene expression regulation.

Introduction—A cell's capacity to thrive in a given environment hinges upon its ability to synthesize mRNAs and proteins at the appropriate levels, neither lacking essential components nor wasting resources on superfluous products [1–5]. This intricate balancing necessitates a regulatory system, the principles governing which remain incompletely understood. In this work, we seek to answer a fundamental question: what are the optimal expression levels of mRNAs and proteins that maximize cell fitness? On the one hand, the protein levels should be above some minimum values to perform their functions appropriately [6], and the mRNA levels also cannot be too low; otherwise, the resulting drastic protein noise may compromise the accuracy of gene expression [7–11]. On the other hand, gene expression also imposes a cost on cell fitness, as excessive expression wastes limited cellular resources and disrupts cellular physiological functions [2, 3, 12–18].

To answer the question, we examine the trade-off between the cost and benefit of mRNA and protein production. We present a mathematical model that incorporates the benefit of protein synthesis, the cost associated with transcription and translation, and the toxicity effects of protein overexpression [3]. Numerous experiments suggested that gene expression levels are under evolutionary selection [2, 19–28]. Therefore, we seek to find the optimal mRNA and protein copy numbers that maximize the average fitness. Our numerical simulations and analytical calculations reveal a scaling relationship between the optimal mRNA copy number (m) and the protein copy number (p): $m \sim p^{\gamma}$. Surprisingly, the γ exponent has a lower bound, 0.5, which occurs when the toxicity effects of protein overexpression are mild; a more substantial toxicity effect results in a larger γ exponent.

To verify our theories, we simulate an *in silico S. cerevisiae* cell using biologically relevant parameters. Intriguingly, the predicted γ exponent is close to 0.5 for the *in silico* cell, and also agrees well with the experimental data. Furthermore, the predicted scaling between m and p is observed in diverse organisms [9, 29–32], including human, mouse, fission yeast, and *E. coli*, where the γ exponent is all above the predicted lower bound. Particularly, *E. coli* exhibits a scaling close to $m \sim p$, suggesting a significant toxicity effect of protein overexpression, consistent with previous experiments [2]. Finally, we reveal the implication of the lower bound of the γ exponent: a γ exponent larger than 0.5 means that the regulation of protein abundance is dominated by transcription, in agreement with various experiments [30, 33–36]. In summary, our work lays an evolutionary foundation for the mRNA-protein correlation across multiple organisms.

Benefit and cost of gene expression—We consider a gene whose encoded protein benefits cell fitness but also imposes a cost. We introduce the benefit function $B(p)$ as the relative increase in fitness due to the produced protein. To simplify without losing the essential ingredient, we set

$$B = (1-b) + b\frac{p^{s_\mathrm{b}}}{p^{s_\mathrm{b}} + p_\mathrm{b}^{s_\mathrm{b}}} \tag{1}$$

where b is a constant. We denote p_b as the minimal protein number needed, and s_b is the sensitivity factor quantifying how sensitive cell fitness is to the protein copy number.

We propose that gene expression cost includes two sources. The first source comes from the global competition between genes for finite resources, including ribosome, RNA polymerase [1, 12, 37–44]. Experiments and theories showed that the gene expression load can be decomposed into a transcription cost (C_tx) and a translation cost (C_tl), each proportional to the mRNA copy number and the protein copy number, respectively (Figure 1a) [1, 3, 45–47]. One should note that the

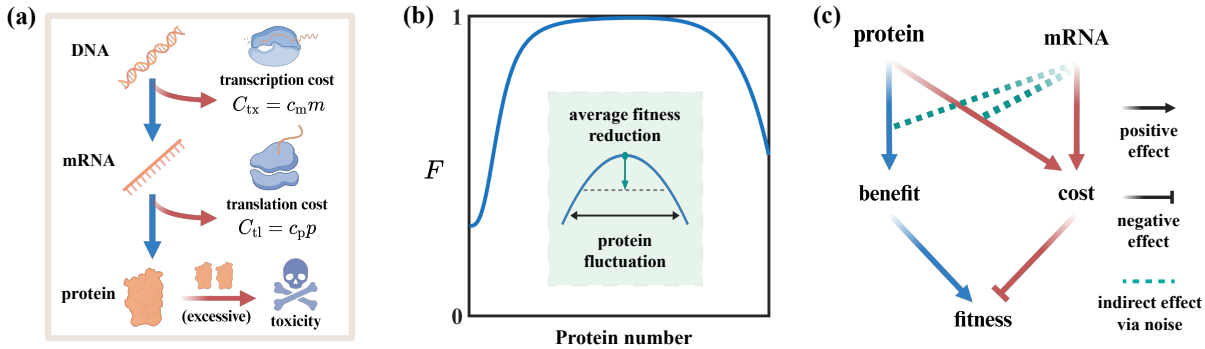

FIG. 1. **The benefit and cost of gene expression determine the fitness.** (a) The cost of gene expression. Gene expression consumes limited cellular resources, imposing a load that reduces cell fitness. The total load lies in both the transcription and the translation processes, each proportional to mRNA and protein copy numbers, respectively. Besides, protein overexpression exacerbates the cost as excessive proteins can be toxic to cells. (b) A schematic of the fitness as a function of protein number. The relative fitness increases from $1-b$, approaches 1, and drops with an overexpression of proteins. The inset shows a reduction in the average fitness due to protein noise. (c) The causal maps of how protein and mRNA influence cell fitness. An insufficiency of mRNA gives rise to protein abundance fluctuation, which indirectly affects the benefit and cost of protein expression.

competition-generated cost is universal and independent of specific protein functions [47, 48].

The other source of cost arises from the toxicity of protein overexpression ($C_{\rm d}$), which we model by a nonlinear function: $C_{\rm d} = (p/p_{\rm d})^{s_{\rm d}}$. Here, $p_{\rm d}$ is a constant which we denote as the toxicity threshold, and $s_{\rm d}$ is the sensitivity factor of cell fitness to protein overexpression. To sum up, the total cost function is

$$C = c_{\rm m} m + c_{\rm p} p + \left(\frac{p}{p_{\rm d}}\right)^{s_{\rm d}}, \tag{2}$$

where $c_{\rm m}$ is the cost per mRNA copy and $c_{\rm p}$ is the cost per protein copy due to resource competition (Figure 1a). Given the cost and benefit, we introduce the total fitness as

$$F(m,p) = B(p) \times [1 - C(m,p)], \tag{3}$$

which is a non-monotonic function of protein abundance (Figure 1b), in agreement with previous experiments [19–21, 23] (Figure S1).

One may expect that the smaller the mRNA copy number, the higher the fitness since it minimizes the transcription cost [Eq. (2)]. However, a small mRNA copy number also leads to a high protein noise, which can reduce the time-averaged or population-averaged fitness (Fig. 1(b)). In this work, we use the following relationship between protein noise and mRNA copy number [49, 50]:

$$\mathrm{CV}^2 = \frac{\tau_{\rm m}}{\tau_{\rm p} m} + \frac{1}{p} + \mathrm{CV}_0^2 \tag{4}$$

where CV is the coefficient of variation of protein number, $\tau_{\rm m}$ is the mRNA lifetime, $\tau_{\rm p}$ is the protein lifetime, and CV_0^2 is the noise floor [7, 49, 51–53]. We summarize the cost and benefit of gene expression, where mRNA and protein affect the fitness through coupled pathways (Figure 1c).

Emerging mRNA-protein scaling—In the following, we find the optimal values of m and p that maximize the average fitness. For a given gene with its mRNA and protein, we generate 10^5 random protein numbers following a Gamma distribution with the mean p and a variance set by the mRNA copy number m according to Eq. (4) [54] and calculate the average fitness $\langle F(m,p)\rangle$. We sample a range of mRNA and protein copy numbers to find the m and p that maximize the average fitness. We change the minimal protein number $p_{\rm b}$ to represent proteins with various cellular demands. For simplicity, we keep other parameters the same for all proteins, as the protein copy number needed for physiological functions accounts for the major differences among genes. Intriguingly, a power-law scaling between the optimal mRNA and protein copy numbers emerges (Figure 2a):

$$m \sim p^{\gamma}. \tag{5}$$

Surprisingly, the γ exponent decreases from a value near $s_d/2$ and approaches a lower bound 0.5 as the toxicity threshold increases (Figure 2b). In the following, we seek to understand the origin of the emerging scaling theoretically by studying a simplified model, and denote the model mentioned so far as the full model.

Lower bound of the scaling exponent—We simplify the full model by assuming a small protein noise and approximate the reduction in the average fitness due to protein noise using the local curvature of the fitness function:

$$\Delta F = -\frac{1}{2}\,|F''(p)|\,p^2 \mathrm{CV}^2 \tag{6}$$

Here, we take p at the peak value of the fitness function

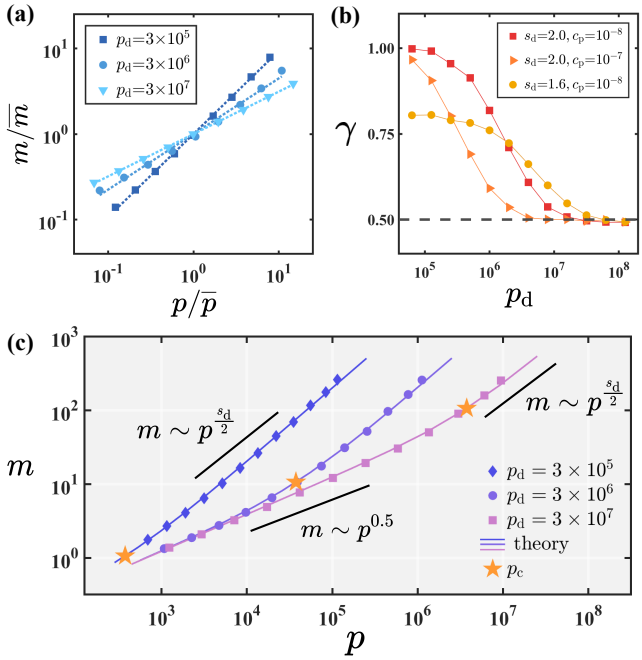

FIG. 2. **The emerging scaling between mRNA and protein copy numbers at optimal fitness.** (a) The mRNA-protein scaling by maximizing the average fitness function considering protein noise, where results with different toxicity thresholds (p_d) are shown. Both mRNA and protein copy numbers are normalized by their mean values. In this figure, we set $\tau_\mathrm{m}/\tau_\mathrm{p} = 0.1$, $\mathrm{CV}^2 = 10^{-2}$, $b = 0.5$, $s_\mathrm{b} = 2$, $s_\mathrm{d} = 2$, $c_\mathrm{m} = 10^{-6}$, $c_\mathrm{p} = 10^{-8}$, and p_b from 1 to 10^4 unless otherwise specified. (b) The scaling exponent γ decreases with p_d, approaching a lower bound 0.5, regardless of other parameters. (c) The analytical predictions Eq. (7) (solid lines) perfectly align with numerical simulations of the full model. When $p \ll p_c$, $m \sim p^{0.5}$, and when $p \gg p_c$, $m \sim p^{s_d/2}$. Here, for each p_d, we scan p_b from 1 to $p_\mathrm{d}/3$.

before considering the noise effects as an approximation; therefore, the first-order derivative does not show up. By further maximizing the fitness with respect to m (End Matter), we find the optimal m value given the optimal p value as

$$m = \sqrt{\alpha\left[1 + \left(\frac{p}{p_\mathrm{c}}\right)^{s_\mathrm{d}-1}\right]} p^{\frac{1}{2}}. \tag{7}$$

Here, α is a constant, and $p_\mathrm{c} \propto p_\mathrm{d}^{s_d/(s_\mathrm{d}-1)}$ is the critical protein number (see their expressions in End Matter). Intriguingly, Eq. (7) shows that $m \sim p^{0.5}$ when the toxicity effect is negligible, i.e., when $p \ll p_c$; it also predicts $m \sim p^{s_\mathrm{d}/2}$ when the toxicity effect is significant, i.e., $p \gg p_c$ (Figure 2c).

Intuitively, when toxicity effects are significant, suppressing the protein noise is critical for the cell to survive; therefore, a higher mRNA copy number is necessary to reduce the noise, generating a higher scaling exponent between m and p. For p that is comparable to p_c, the m vs. p curve exhibits an apparent scaling exponent $0.5 < \gamma < s_\mathrm{d}/2$ (Figure 2c). Indeed, the $m \sim p^\gamma$ scalings from the simulations of the full model (Figure 2a) sit in the different regimes of the analytical prediction of the simplified model (Figure 2c), strongly supporting the validity of the simplified model.

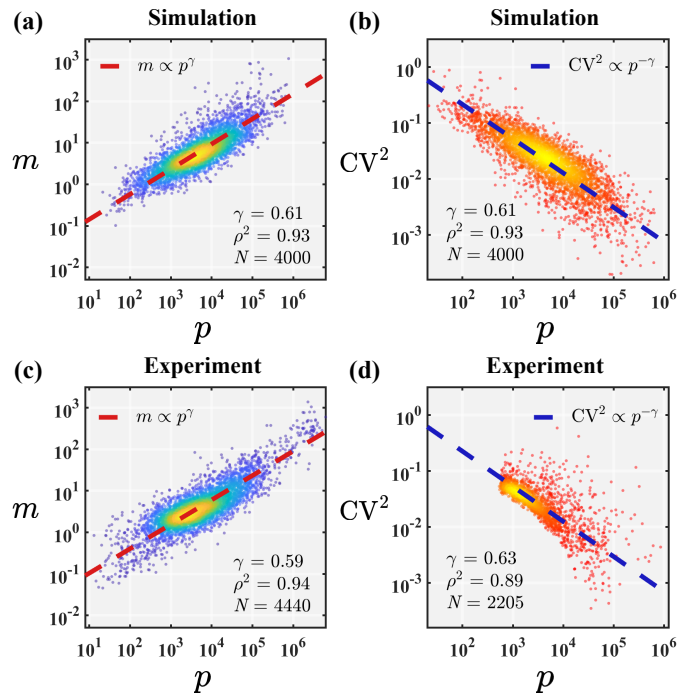

FIG. 3. **Comparison between the model predictions and experimental data of** *S. cerevisiae***.** (a-b) The mRNA copy number and the CV^2 of protein copy number vs. the protein copy number for the *in silico* cell under the conditions of maximized averaged fitness. The numerical methods are the same as Figure 2. The color of the dots, yellow to blue for (a), red to orange for (b), represents the local density, from high to low. (c-d) The mRNA copy number and the CV^2 of protein copy number vs. the protein copy number for the experimental data from Ref. [29] and [9], respectively. To clearly show the scaling of CV^2 vs. p, we remove the noise floor $\mathrm{CV}_0^2 = 10^{-2}$ in the plots. The dashed lines (a-d) are power-law fittings. ρ^2 is the fraction of total variance explained by the power-law scaling $m \sim p^\gamma$, and N is the number of genes detected or simulated. The γ exponent is calculated by the total-least-squares (TLS) method (Table S2 and Figure S4), which is also used in other relevant figures unless otherwise mentioned.

In silico cells vs. experiments—In the following, we simulate an *in silico* cell using biologically relevant parameters of *S. cerevisiae* (see details in Supplemental Material). In particular, we estimate s_b, p_b, b, s_d, and p_d by fitting the experimental measured fitness function vs. protein abundance [19]. We also estimate the cost coefficients c_m and c_p based on resource competition [48], supported by experimental measurements [3]. For each gene, we randomly sample parameters for the *in silico*

cell from distributions that make good approximations of the experimental data (Table S1 and Figure S2), and compute the optimal mRNA and protein copy numbers, and also the CV^2 of the protein copy numbers.

We notice that toxicity effects due to protein overexpression are generally mild in *S. cerevisiae* (Figure S1), suggesting that $p \ll p_c$ for most genes; therefore, we expect the scaling exponent between m and p to be close to 0.5. Indeed, the optimal mRNA and protein copy number exhibits a scaling $m \sim p^\gamma$ where $\gamma \approx 0.6$ (Figure 3a). The CV^2 of protein number also scales as $\mathrm{CV}^2 \sim p^{-\gamma}$ as predicted by Eq. (4), where the upstream noise dominates the protein noise due to a limited mRNA copy number (Figure 3b).

Surprisingly, the experimental data match our theoretical predictions regarding both the scaling exponent and the absolute values (Figure 3c, d), which strongly supports the validity of our model: balancing the cost and benefit of gene expression is the driving force of gene expression regulation. Meanwhile, the *in silico* simulation demonstrates that our model is robust against genetic heterogeneity since all the genes have different parameters. For *S. cerevisiae*, translational regulation is important because $p/m \sim m^{0.7}$: the higher mRNA copy numbers, the more protein copies produced per mRNA copy.

mRNA-protein scaling across different organisms—We further analyze the genome-wide data of multiple organisms, including *E. coli* [30, 50, 56], *S. cerevisiae* [29, 57], *Schizosaccharomyces pombe* (fission yeast) [31], *Mus musculus* (mouse) [32, 55], and *Homo sapiens* (human) [32, 58]. Intriguingly, all the measured exponents between m and p are above 0.5 (Figure 4a-d and Figure S3), in perfect agreement with our model predictions. We remark that the lower bound of the γ exponent is a consequence of fitness optimization, and no biochemical constraints require that γ has to be larger than 0.5.

What does the lower bound 0.5 mean for the γ exponent? Biologically, the expression level of a given protein depends on transcriptional regulation and translational regulation; therefore, one can decompose the protein copy number into two factors $p = p_1 p_2$ where $p_1 \sim m$ is the expected protein level if only transcriptional regulation occurs and p_2 represents the extra modulation due to translational regulation. From $m \sim p^\gamma$, it is straightforward to find that $p_2/p_1 \sim m^\delta$ where $\delta \equiv \frac{1}{\gamma} - 2 < 0$. A negative δ exponent means that transcriptional regulation dominates over translational regulation, especially for highly expressed genes. Intriguingly, for the model organism *E. coli*, we find a γ exponent near 1, suggesting that translational regulation is essentially absent [30].

Discussion—Optimizing fitness under a cost-benefit trade-off has been applied to determine the expression levels of particular genes both theoretically and experimentally [2, 56, 59, 60]. However, whether a genome-wide scaling between the mRNA and protein expression levels exists due to cost-benefit trade-offs is still elusive. In this work, we unveil a power-law scaling between the mRNA and protein copy numbers that maximizes cell fitness, and predict a lower bound for the γ exponent, which is 0.5. We further show that the γ exponent depends on the toxicity threshold of protein copy number, and a lower toxicity threshold leads to a higher exponent.

By simulating an *in silico* cell using biologically relevant parameters of *S. cerevisiae*, we demonstrate that our model quantitatively aligns with experiments. In particular, the γ exponent for *S. cerevisiae* is near the lower bound 0.5, suggesting a mild toxicity effect of protein overexpression. We further compute the γ exponents for

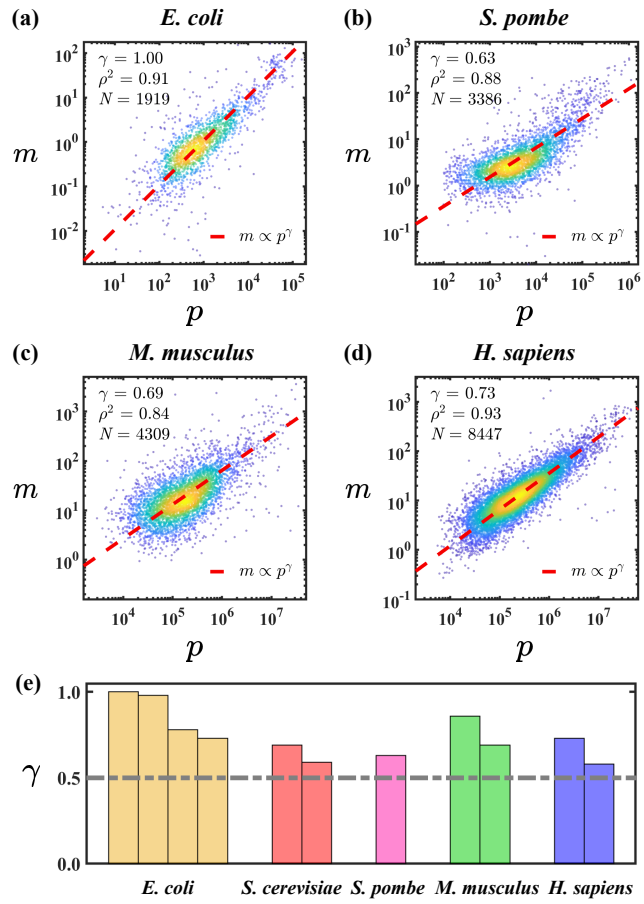

FIG. 4. **The m-p scaling across multiple organisms.** (a-d) The experimental data of m vs. p for multiple organisms. The color of the dots (yellow to blue) represents the local density (high to low). Data resources are as follows: *E. coli* [30], *S. pombe* [31], *M. musculus* (NIH 3T3 cell) [55] (the bias in protein measurement corrected by [33]), and *H. sapiens* (HeLa cell) [32]. The color of the dots (yellow to blue) represents the local density (high to low). ρ^2 is the fraction of total variance explained by the power-law scaling $m \sim p^\gamma$, and N is the number of genes detected. (e) The scaling exponents in different species are all higher than 0.5. Each bar represents one dataset, where the original m-p relations are shown in (a-d) and Figure S3.

multiple organisms and find that in all cases, the exponent is above the predicted lower bound. According to our theory, the different γ exponents originate from the distinct toxicity effects of protein overexpression. For *S. cerevisiae*, the expression levels of most proteins are far below the critical protein number ($p \ll p_c$) such that protein toxicity is neglible; in contrast, for *E. coli*, the toxicity effects are likely to happen near the expression levels of wild-type cells.

We also unveil the implications for the lower bound: protein expression levels are primarily set by mRNA expression levels, in particular for highly expressed genes, in agreement with numerous experiments [30, 33–35, 61–66]. Our theory does not exclude the possibility of a γ exponent larger than 1, and it will be interesting to see whether this situation can happen in some organisms.

Acknowledgments—The research was funded by the National Key Research and Development Program of China (2024YFA0919600) and supported by Peking-Tsinghua Center for Life Sciences grants.

[1] A. Wagner, Energy constraints on the evolution of gene expression, Molecular Biology and Evolution **22**, 1365 (2005).

[2] E. Dekel and U. Alon, Optimality and evolutionary tuning of the expression level of a protein, Nature **436**, 588 (2005).

[3] M. Kafri, E. Metzl-Raz, G. Jona, and N. Barkai, The cost of protein production, Cell reports **14**, 22 (2016).

[4] J. M. Peters, A. Colavin, H. Shi, T. L. Czarny, M. H. Larson, S. Wong, J. S. Hawkins, C. H. Lu, B.-M. Koo, E. Marta, *et al.*, A comprehensive, crispr-based functional analysis of essential genes in bacteria, Cell **165**, 1493 (2016).

[5] T. W. Lo, H. J. Choi, D. Huang, and P. A. Wiggins, Noise robustness and metabolic load determine the principles of central dogma regulation, Science Advances **10**, eado3095 (2024).

[6] N. M. Belliveau, G. Chure, C. L. Hueschen, H. G. Garcia, J. Kondev, D. S. Fisher, J. A. Theriot, and R. Phillips, Fundamental limits on the rate of bacterial growth and their influence on proteomic composition, Cell Systems **12**, 924 (2021).

[7] J. Paulsson, Summing up the noise in gene networks, Nature **427**, 415 (2004).

[8] J. M. Raser and E. K. O'shea, Noise in gene expression: origins, consequences, and control, Science **309**, 2010 (2005).

[9] J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise, Nature **441**, 840 (2006).

[10] A. Sanchez, S. Choubey, and J. Kondev, Regulation of noise in gene expression, Annual review of biophysics **42**, 469 (2013).

[11] J. Lin and A. Amir, The effects of stochasticity at the single-cell level and cell size control on the population growth, Cell Systems **5**, 358 (2017).

[12] W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis, and D. S. Kompala, Plasmid-encoded protein: the principal factor in the metabolic burden associated with recombinant bacteria, Biotechnology and bioengineering **35**, 668 (1990).

[13] H. Dong, L. Nilsson, and C. G. Kurland, Gratuitous overexpression of genes in escherichia coli leads to growth inhibition and ribosome destruction, Journal of Bacteriology **177**, 1497 (1995).

[14] C. Rang, J. E. Galen, J. B. Kaper, and L. Chao, Fitness cost of the green fluorescent protein in gastrointestinal bacteria, Canadian journal of microbiology **49**, 531 (2003).

[15] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa, Interdependence of cell growth and gene expression: origins and consequences, Science **330**, 1099 (2010).

[16] M. Lynch and G. K. Marinov, The bioenergetic costs of a gene, Proceedings of the National Academy of Sciences **112**, 15690 (2015).

[17] A. San Millan, M. Toll-Riera, Q. Qi, and R. C. MacLean, Interactions between horizontally acquired genes create a fitness cost in pseudomonas aeruginosa, Nature communications **6**, 6845 (2015).

[18] Y. Eguchi, K. Makanae, T. Hasunuma, Y. Ishibashi, K. Kito, and H. Moriya, Estimating the protein burden limit of yeast cells by measuring the expression limits of glycolytic proteins, eLife **7**, e34595 (2018).

[19] L. Keren, J. Hausser, M. Lotan-Pompan, I. V. Slutskin, H. Alisar, S. Kaminski, A. Weinberger, U. Alon, R. Milo, and E. Segal, Massively parallel interrogation of the effects of gene expression levels on fitness, Cell **166**, 1282 (2016).

[20] J. S. Rest, C. M. Morales, J. B. Waldron, D. A. Opulente, J. Fisher, S. Moon, K. Bullaughey, L. B. Carey, and D. Dedousis, Nonlinear fitness consequences of variation in expression level of a eukaryotic gene, Molecular Biology and Evolution **30**, 448 (2012).

[21] F. Duveau, W. Toubiana, and P. J. Wittkopp, Fitness effects of cis-regulatory variants in the saccharomyces cerevisiae tdh3 promoter, Molecular Biology and Evolution **34**, 2908 (2017).

[22] F. Duveau, A. Hodgins-Davis, B. P. Metzger, B. Yang, S. Tryban, E. A. Walker, T. Lybrook, and P. J. Wittkopp, Fitness effects of altering gene expression noise in saccharomyces cerevisiae, eLife **7**, e37272 (2018).

[23] F. Duveau, P. Vande Zande, B. P. Metzger, C. J. Diaz, E. A. Walker, S. Tryban, M. A. Siddiq, B. Yang, and P. J. Wittkopp, Mutational sources of trans-regulatory variation affecting gene expression in saccharomyces cerevisiae, eLife **10**, e67806 (2021).

[24] G. I. Lang, A. W. Murray, and D. Botstein, The cost of gene expression underlies a fitness trade-off in yeast, Proceedings of the National Academy of Sciences **106**, 5755 (2009).

[25] F. J. Poelwijk, P. D. Heyning, M. G. de Vos, D. J. Kiviet, and S. J. Tans, Optimality and evolution of transcriptionally regulated gene expression, BMC systems biology **5**, 128 (2011).

[26] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain, Mechanistic links between cellular trade-offs, gene expression, and growth, Proceedings of the National Academy of Sciences **112**, E1038 (2015).

[27] P. Vande Zande, M. S. Hill, and P. J. Wittkopp,

Pleiotropic effects of trans-regulatory mutations on fitness and gene expression, Science **377**, 105 (2022).
[28] A. Gyorgy, Competition and evolutionary selection among core regulatory motifs in gene expression control, Nature communications **14**, 8266 (2023).
[29] D. E. Weinberg, P. Shah, S. W. Eichhorn, J. A. Hussmann, J. B. Plotkin, and D. P. Bartel, Improved ribosome-footprint and mrna measurements provide insights into dynamics and regulation of yeast translation, Cell reports **14**, 1787 (2016).
[30] R. Balakrishnan, M. Mori, I. Segota, Z. Zhang, R. Aebersold, C. Ludwig, and T. Hwa, Principles of gene regulation quantitatively connect dna to rna and proteins in bacteria, Science **378**, eabk2066 (2022).
[31] S. Marguerat, A. Schmidt, S. Codlin, W. Chen, R. Aebersold, and J. Bähler, Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells, Cell **151**, 671 (2012).
[32] S. W. Eichhorn, H. Guo, S. E. McGeary, R. A. Rodriguez-Mias, C. Shin, D. Baek, S.-h. Hsu, K. Ghoshal, J. Villén, and D. P. Bartel, mrna destabilization is the dominant effect of mammalian micrornas by the time substantial repression ensues, Molecular Cell **56**, 104 (2014).
[33] J. J. Li, P. J. Bickel, and M. D. Biggin, System wide analyses have underestimated protein abundances and the importance of transcription in mammals, PeerJ **2**, e270 (2014).
[34] J. J. Li and M. D. Biggin, Statistics requantitates the central dogma, Science **347**, 1066 (2015).
[35] Y. Liu, A. Beyer, and R. Aebersold, On the dependency of cellular protein levels on mrna abundance, Cell **165**, 535 (2016).
[36] C. Buccitelli and M. Selbach, mrnas, proteins and the emerging principles of gene expression control, Nature Reviews Genetics **21**, 630 (2020).
[37] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, and J. B. Plotkin, Rate-limiting steps in yeast protein translation, Cell **153**, 1589 (2013).
[38] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, Quantifying cellular capacity identifies gene expression designs with reduced burden, Nature methods **12**, 415 (2015).
[39] G. Wu, Q. Yan, J. A. Jones, Y. J. Tang, S. S. Fong, and M. A. Koffas, Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications, Trends in Biotechnology **34**, 652 (2016).
[40] J. Lin and A. Amir, Homeostasis of protein and mrna concentrations in growing cells, Nature communications **9**, 4496 (2018).
[41] Y. Han and F. Zhang, Control strategies to manage trade-offs during microbial production, Current Opinion in Biotechnology **66**, 158 (2020).
[42] Q. Wang and J. Lin, Heterogeneous recruitment abilities to rna polymerases generate nonlinear scaling of gene expression with cell volume, Nature communications **12**, 6852 (2021).
[43] M. P. Swaffer, G. K. Marinov, H. Zheng, L. F. Valenzuela, C. Y. Tsui, A. W. Jones, J. Greenwood, A. Kundaje, W. J. Greenleaf, R. Reyes-Lamothe, *et al.*, Rna polymerase ii dynamics and mrna stability feedback scale mrna amounts with cell size, Cell **186**, 5254 (2023).
[44] L. Calabrese, L. Ciandrini, and M. Cosentino Lagomarsino, How total mrna influences cell growth, Proceedings of the National Academy of Sciences **121**, e2400679121 (2024).
[45] I. Frumkin, D. Schirman, A. Rotman, F. Li, L. Zahavi, E. Mordret, O. Asraf, S. Wu, S. F. Levy, and Y. Pilpel, Gene architectures that minimize cost of gene expression, Molecular Cell **65**, 142 (2017).
[46] K. Tomala and R. Korona, Evaluating the fitness cost of protein expression in saccharomyces cerevisiae, Genome Biology and Evolution **5**, 2051 (2013).
[47] D. M. Stoebel, A. M. Dean, and D. E. Dykhuizen, The cost of expression of escherichia coli lac operon proteins is in the process, not in the products, Genetics **178**, 1653 (2008).
[48] Y. Yan and J. Lin, Disentangling the cost of gene expression, bioRxiv , 2025 (2025).
[49] J. Paulsson, Models of stochastic gene expression, Physics of life reviews **2**, 157 (2005).
[50] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells, Science **329**, 533 (2010).
[51] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, Noise in protein expression scales with natural protein abundance, Nature genetics **38**, 636 (2006).
[52] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, Proceedings of the National Academy of Sciences **105**, 17256 (2008).
[53] M. Mundt, A. Anders, S. M. Murray, and V. Sourjik, A system for gene expression noise control in yeast, ACS synthetic biology **7**, 2618 (2018).
[54] N. Friedman, L. Cai, and X. S. Xie, Linking stochastic dynamics to population distribution: An analytical framework of gene expression, Phys. Rev. Lett. **97**, 168302 (2006).
[55] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, Global quantification of mammalian gene expression control, Nature **473**, 337 (2011).
[56] G.-W. Li, D. Burkhardt, C. Gross, and J. S. Weissman, Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources, Cell **157**, 624 (2014).
[57] F. W. Albert, D. Muzzey, J. S. Weissman, and L. Kruglyak, Genetic influences on translation in yeast, PLoS genetics **10**, e1004692 (2014).
[58] N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, and M. Mann, Deep proteome and transcriptome mapping of a human cancer cell line, Molecular systems biology **7**, 548 (2011).
[59] T. Kalisky, E. Dekel, and U. Alon, Cost–benefit theory and optimal design of gene regulation functions, Physical Biology **4**, 229 (2007).
[60] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa, Emergence of robust growth laws from optimal regulation of ribosome synthesis, Molecular systems biology **10**, 747 (2014).
[61] P. Lu, C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, Nature biotechnology **25**, 117 (2007).
[62] E. Lundberg, L. Fagerberg, D. Klevebring, I. Matic, T. Geiger, J. Cox, C. Älgenäs, J. Lundeberg, M. Mann, and M. Uhlen, Defining the transcriptome and proteome in three functionally different human cell lines, Molecular

systems biology **6**, 450 (2010).
[63] M. Jovanovic, M. S. Rooney, P. Mertins, D. Przybylski, N. Chevrier, R. Satija, E. H. Rodriguez, A. P. Fields, S. Schwartz, R. Raychowdhury, *et al.*, Dynamic profiling of the protein life cycle in response to pathogens, Science **347**, 1259038 (2015).
[64] A. M. Franks, G. Csárdi, D. A. Drummond, and E. M. Airoldi, Estimating a structured covariance matrix from multilab measurements in high-throughput biology, Journal of the American Statistical Association **110**, 27 (2015).
[65] A. Koussounadis, S. P. Langdon, I. H. Um, D. J. Harrison, and V. A. Smith, Relationship between differentially expressed mrna and mrna-protein correlations in a xenograft model system, Scientific reports **5**, 10775 (2015).
[66] D. Choe, J. H. Lee, M. Yoo, S. Hwang, B. H. Sung, S. Cho, B. Palsson, S. C. Kim, and B.-K. Cho, Adaptive laboratory evolution of a genome-reduced escherichia coli, Nature communications **10**, 935 (2019).

End Matter

We first find the optimal p that maximizes the fitness before considering the noise effects, where

$$\begin{aligned} F &= \left[1 - b\frac{p_{\mathrm{b}}^{s_{\mathrm{b}}}}{p^{s_{\mathrm{b}}} + p_{\mathrm{b}}^{s_{\mathrm{b}}}}\right]\left[1 - c_{\mathrm{m}}m - c_{\mathrm{p}}p - \left(\frac{p}{p_{\mathrm{d}}}\right)^{s_{\mathrm{d}}}\right] \\ &\approx 1 - b\left(\frac{p}{p_{\mathrm{b}}}\right)^{-s_{\mathrm{b}}} - c_{\mathrm{m}}m - c_{\mathrm{p}}p - \left(\frac{p}{p_{\mathrm{d}}}\right)^{s_{\mathrm{d}}}. \end{aligned} \tag{8}$$

Here, we use the approximation that the reduction in the benefit function relative to its maximum value 1 due to protein deficiency and the cost are both small around the optimal p; therefore, higher-order terms are neglected, and the optimal p value should be much larger than the minimal protein number p_{b} [5]. The agreement between the theoretical predictions and the numerical calculations of the full model supports the validity of these approximations. We compute $\partial F(m,p)/\partial p = 0$ and find that

$$bs_{\mathrm{b}}\left(\frac{p_{\mathrm{b}}}{p}\right)^{s_{\mathrm{b}}} - s_{\mathrm{d}}\left(\frac{p}{p_{\mathrm{d}}}\right)^{s_{\mathrm{d}}} - c_{\mathrm{p}}p = 0 \tag{9}$$

We further calculate the second derivative of the fitness function at the optimal p value,

$$F''(p) = -\frac{c_{\mathrm{p}}(s_{\mathrm{b}}+1)}{p}\left[1 + \frac{(s_{\mathrm{b}}+s_{\mathrm{d}})s_{\mathrm{d}}}{(s_{\mathrm{b}}+1)c_{\mathrm{p}}p}\left(\frac{p}{p_{\mathrm{d}}}\right)^{s_{\mathrm{d}}}\right]. \tag{10}$$

We next find the optimal m value by computing $\partial\langle F(m,p)\rangle/\partial m = 0$ where

$$\langle F(m,p)\rangle = F(m,p) - \frac{1}{2}|F''(p)|p^2\mathrm{CV}^2, \tag{11}$$

and find that

$$m = \sqrt{\frac{\tau_{\mathrm{m}}|F''(p)|}{2\tau_{\mathrm{p}}c_{\mathrm{m}}}}p = \sqrt{\alpha\left[1 + \left(\frac{p}{p_{\mathrm{c}}}\right)^{s_{\mathrm{d}}-1}\right]}p^{\frac{1}{2}}. \tag{12}$$

Here, $\alpha = \frac{c_{\mathrm{p}}\tau_{\mathrm{m}}(s_{\mathrm{b}}+1)}{2c_{\mathrm{m}}\tau_{\mathrm{p}}}$, and $p_{\mathrm{c}} = \left[\frac{(s_{\mathrm{b}}+1)c_{\mathrm{p}}}{(s_{\mathrm{b}}+s_{\mathrm{d}})s_{\mathrm{d}}}p_{\mathrm{d}}^{s_{\mathrm{d}}}\right]^{\frac{1}{s_{\mathrm{d}}-1}}$.