

Fitness-Driven Scaling Laws between mRNA and Protein Levels

Yichen Yan (颜逸辰)¹ and Jie Lin (林杰)^{1,2,*}

¹Peking-Tsinghua Center for Life Sciences, *Peking University*, Beijing, China

²Center for Quantitative Biology, *Peking University*, Beijing, China



(Received 29 September 2025; accepted 12 June 2026; published 22 July 2026)

The relationship between protein and messenger RNA (mRNA) levels is a key aspect of gene expression. Although a strong mRNA-protein correlation is widely observed, the evolutionary force driving the correlation remains poorly understood, undermining the reliability of using mRNA abundance as a proxy for protein levels. Here, by balancing the cost and benefit of gene expression, we uncover an optimal scaling between mRNA (m) and protein (p) levels that maximizes cell fitness: $m \sim p^\gamma$. Notably, we predict a lower bound of 0.5 for the exponent, and the toxicity effects of protein overexpression further increase it. For the model organism *Saccharomyces cerevisiae*, the simulated scaling across the genome, incorporating biologically relevant parameters, aligns well with experimental data. Genome-wide data across multiple organisms all exhibit a scaling exponent above the predicted lower bound. Our Letter rationalizes the primary role of transcriptional regulation and also the significance of translational regulation in determining protein levels from an evolutionary perspective.

DOI: 10.1103/gkzp-r681

Introduction—A cell's capacity to thrive in a given environment hinges upon its ability to synthesize messenger RNAs (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 Letter, 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 toxic 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 levels that maximize the fitness. Our numerical simulations and analytical calculations reveal a scaling relationship between the optimal mRNA copy number (m) and the protein copy number (p): $m \sim p^\gamma$, which can be for multiple genes across the genome or for a single gene across different conditions. 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. Using biologically relevant parameters, we simulate an *in silico* *S. cerevisiae* cell and the resulting mRNA-protein scaling across the genome agrees well with the experimental data.

Significantly, 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 exponents are all above the predicted lower bound. Critically, we reveal the implication of the lower bound of the γ exponent: protein abundance is primarily determined by transcription, with translation also potentially playing a significant role—a conclusion consistent with multiple experimental findings [30,32–39]. 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. In this Letter, we define the fitness of a gene as the cellular growth rate normalized by the maximal growth rate, $F = \lambda/\lambda_{\max}$ where $\lambda_{\max}$ is attained if the protein level of the particular gene approaches infinity

*Contact author: linjie@pku.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](#). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

and its expression is costless, given that the expression levels of all other genes are fixed. To define the benefit of a protein, we introduce λ_0 as the hypothetical growth rate assuming the expression of the corresponding gene is costless, and introduce the benefit function as $B(p) = \lambda_0(p)/\lambda_{\max}$, which has the following form:

$$B = (1 - b) + b \frac{p^{s_b}}{p^{s_b} + p_b^{s_b}}, \quad (1)$$

where b is a constant. We denote p_b as the minimal protein level needed, and s_b as the sensitivity factor quantifying how sensitive fitness is to the protein level. We note that the Hill-function form of Eq. (1) accurately describes the dependence of fitness on protein levels (Fig. S1). Also, the detailed form of the benefit function does not affect our main conclusions (Sec. A and Fig. S3 in the Supplemental Material [40]).

We define the cost function as the relative reduction of growth rate, $C = 1 - \lambda/\lambda_0$, where λ_0 is the hypothetical growth rate without gene expression cost. We propose that gene expression cost includes two sources. The first is the production cost, which comes from the global competition between genes for limited resources, including adenosine triphosphate, ribosome, ribonucleic acid (RNA) polymerase, etc. [1,74–81]. The expression of a specific gene inevitably consumes these limited resources, thereby suppressing the expression of other genes and ultimately decreasing the growth rate [2,3,12–15]. Experiments and theories reveal that the production cost can be decomposed into a transcription cost (C_{tx}) and a translation cost (C_{tl}), each proportional to the mRNA and protein levels, respectively [Fig. 1(a)] [1,3,16,82–84]. One should note that the production cost is universal and independent of specific protein functions.

The other source of cost arises from the toxicity of protein overexpression, denoted as C_d . We model C_d as

$(p/p_d)^{s_d}$, which nicely captures the nonlinear effect observed in experiments (Figs. S1 and S2). Here, p_d is a gene-specific constant which we denote as the toxicity threshold, and s_d is the sensitivity factor of cell fitness to protein overexpression. We remark that proteins that exhibit little toxicity upon overexpression have a large toxicity threshold p_d , which means a small C_d . Also, the detailed form of the toxicity function does not affect our main conclusions (Sec. A and Fig. S3 in the Supplemental Material [40]).

To sum up, the total cost function is

$$C = c_m m + c_p p + \left(\frac{p}{p_d}\right)^{s_d}, \quad (2)$$

where c_m is the cost per mRNA copy and c_p is the cost per protein copy due to resource competition [Fig. 1(a)]. Given that the maximum relative growth rate reduction is 1 (at growth arrest), C is assigned the minimum between 1 and the value from Eq. (2). Given the cost and benefit, we introduce the total fitness as

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

which is a nonmonotonic function of protein abundance [Fig. 1(b)], in agreement with previous experiments [19–22] (Fig. S1). Since both B and C are between 0 and 1, the fitness is always non-negative.

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 Letter, we use the following relationship between protein noise and mRNA level [85,86]:

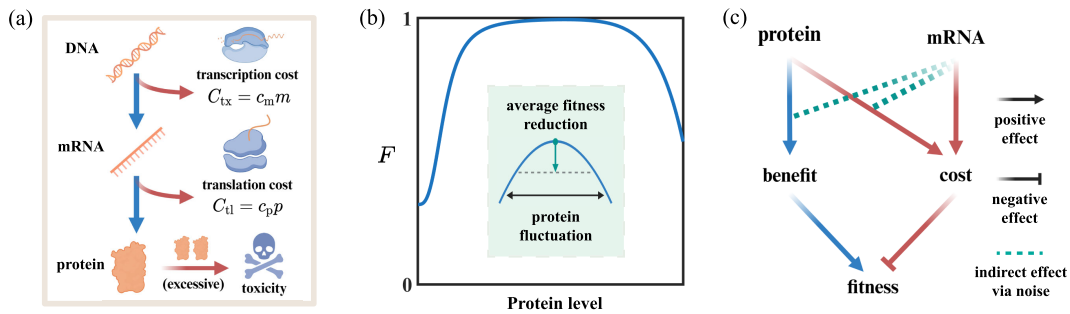


FIG. 1. The benefit and cost of gene expression determine the fitness. (a) The cost of gene expression includes the production cost and the protein toxicity. The production cost lies in the transcription and translation processes, each proportional to mRNA and protein levels, respectively. Besides, protein overexpression exacerbates the cost as excessive proteins can be toxic to cells. (b) A schematic of the fitness as a function of protein level. The relative fitness increases from $1 - b$, approaches 1, and then drops as the protein is overexpressed. 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.

$$CV^2 = \frac{\tau_m}{\tau_p m} + \frac{1}{p} + CV_0^2, \quad (4)$$

where CV is the coefficient of variation of protein level, τ_m is the mRNA lifetime, τ_p is the protein lifetime, and CV_0 is the noise floor [7,85,87–89]. We summarize the cost and benefit of gene expression, where mRNA and protein affect the fitness through coupled pathways [Fig. 1(c)].

Emerging mRNA-protein scaling—In the following, we find the optimal values of m and p that maximize the average fitness. The mRNA and protein levels are treated as independent variables in the optimization problem. 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) [90] and calculate the average fitness $\langle F(m, p) \rangle$. We sample a range of mRNA and protein levels to find the m and p that maximize the average fitness. We change the minimal protein level p_b to represent proteins with various cellular demands. For simplicity, we keep other parameters constant, since the protein level required for physiological functions accounts for the major differences among genes or conditions. Intriguingly, a power-law scaling between the optimal mRNA and protein levels emerges [Fig. 2(a)]:

$$m \sim p^\gamma. \quad (5)$$

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

Lower bound of the scaling exponent—We simplify the full model by assuming a small protein noise and first find the optimal p by maximizing the fitness $F(m, p)$ without considering noise effects. We then approximate the reduction in the average fitness due to protein noise using the local curvature of the fitness function:

$$\Delta F = -\frac{1}{2} \left| \frac{\partial^2 F}{\partial p^2} \right| p^2 CV^2. \quad (6)$$

Here, the first-order derivative does not show up because p is at the peak value of the fitness function. By further maximizing the fitness with respect to m (see the End Matter), we find the optimal m value given the optimal p value as

$$m = \sqrt{\alpha \left[1 + \left(\frac{p}{p_c} \right)^{s_d-1} \right]} p^{\frac{1}{2}}. \quad (7)$$

Here, α is a constant, and $p_c \propto p_d^{s_d/(s_d-1)}$ is the critical protein level (see their expressions in the End Matter). Intriguingly,

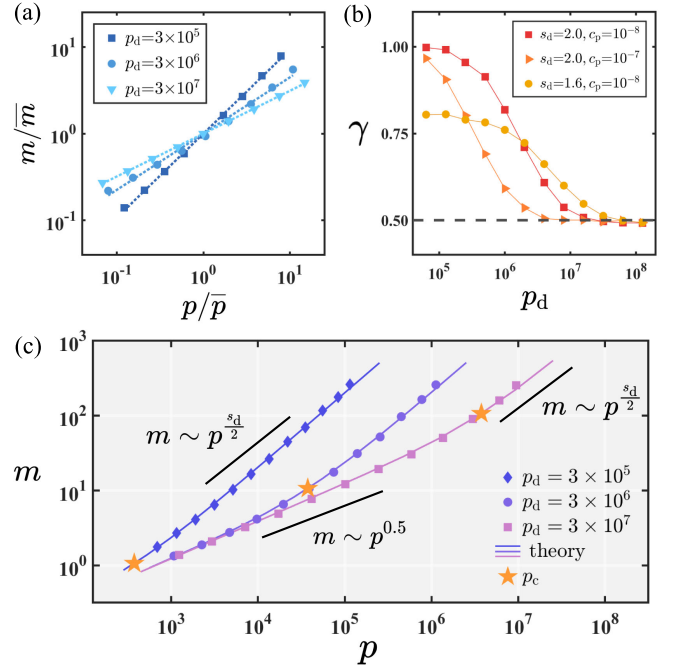


FIG. 2. The emerging scaling between mRNA and protein levels at optimal fitness. (a) The mRNA-protein scaling by maximizing the average fitness function considering protein noise, where results with different toxicity thresholds (p_d) are shown. A larger p_d means lower toxicity. Here, mRNA and protein levels are normalized by their mean values. In this figure, we set $\tau_m/\tau_p = 0.1$, $CV_0^2 = 10^{-2}$, $b = 0.5$, $s_b = 2$, $s_d = 2$, $c_m = 10^{-6}$, $c_p = 10^{-8}$, 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 from Eq. (7) (solid lines) match the numerical simulations of the full model perfectly. When $p \ll p_c$ (negligible toxicity), $m \sim p^{0.5}$, and when $p \gg p_c$ (strong toxicity), $m \sim p^{s_d/2}$. Here, for each p_d , we scan p_b from 1 to $p_d/3$.

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 $m \sim p^{s_d/2}$ when the toxicity effect is significant, i.e., $p \gg p_c$, in agreement with our numerical simulations [Fig. 2(b)]. The predicted m - p relation also quantitatively matches the simulation results of the full model [Fig. 2(c)]. Intuitively, when toxicity effects are significant, suppressing protein noise is critical for the cell to survive; therefore, a higher mRNA level is necessary to reduce the noise, leading to 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 $0.5 < \gamma < s_d/2$ [Fig. 2(c)]. We highlight that the m - p scaling is both applicable to a group of genes under the same condition, as we show later, or a single gene under different conditions (Sec. B and Figs. S4 and S5 in the Supplemental Material [40]).

We note that the lower bound of the γ exponent suggests that transcriptional regulation plays a primary role in setting protein levels. To see this, we consider the fold change in the expression levels of two proteins under the same

condition (or the fold change in the expression level of the same protein under two conditions). Using the scaling relation, $m \sim p^\gamma$, we find that

$$\log\left(\frac{p_2}{p_1}\right) = \log\left(\frac{m_2}{m_1}\right) + \frac{1-\gamma}{\gamma} \log\left(\frac{m_2}{m_1}\right). \quad (8)$$

Here, the first term on the right side of Eq. (8) represents the fold change due to transcriptional regulation; the second term represents the fold change due to translational regulation. Because $\gamma > 0.5$ [i.e., $(1-\gamma)/\gamma < 1$], the contribution from transcriptional regulation in determining protein levels is larger than that from translational regulation.

We also extend our model to the case in which gene interactions exist, such that a signaling protein regulates the expression levels of other genes. We find that the mRNA-protein scaling still holds across multiple genes under the same conditions, and across different conditions for a single gene. Importantly, our conclusion regarding the bound on the scaling exponent between mRNA and protein levels remains unchanged, even in the presence of strong correlations in protein levels across genes (Sec. B and Figs. S4 and S5 in the Supplemental Material [40]).

In silico cells vs experiments—In the following, we simulate the mRNA-protein relationship across the genome for an *in silico* cell using biologically relevant parameters of *S. cerevisiae* (see details in Sec. C of the Supplemental Material [40]). 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 [84], 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, Fig. S6, and Sec. C of the Supplemental Material [40]), 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* (Fig. 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 levels exhibit a scaling around $m \sim p^{0.61}$ [Fig. 3(a)]. The CV^2 of the protein number also scales as $CV^2 \sim p^{-\gamma}$ as predicted by Eq. (4), where the upstream noise dominates the protein noise due to a limited mRNA level [Fig. 3(b)].

Surprisingly, the experimental data match our theoretical predictions for both the scaling exponent and the absolute values [Figs. 3(c) and 3(d)], strongly supporting the validity of our model given that the simulation parameters are estimated from independent experiments. The *in silico* simulations also demonstrate that our model is robust to genetic heterogeneity, since the genes have different

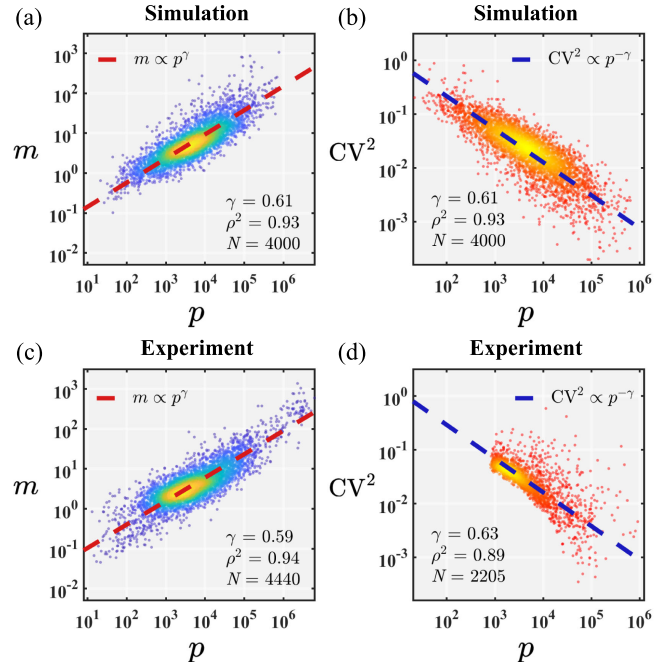


FIG. 3. Comparison between the model predictions and experimental data of *S. cerevisiae*. (a),(b) The mRNA level and the CV^2 of protein level vs the protein level for the *in silico* cell under the conditions of maximized averaged fitness. The numerical methods are the same as Fig. 2. (c),(d) Experimental m vs p [29] and CV^2 vs p [9], respectively. The color of the dots, yellow to blue for (a), yellow to red for (b), represents the local density from high to low. 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. To clearly show the scaling of CV^2 vs p , we remove the noise floor term $CV_0^2 = 10^{-2}$ in the plots. The γ exponent is calculated by the total-least-squares method (Table S2, Fig. S8, and Sec. D in the Supplemental Material [40]), which is also used in other relevant figures unless otherwise mentioned.

parameters. We note that another independent dataset also matches our predictions reasonably well [Fig. S7(d)], further validating our model.

mRNA-protein scaling across different organisms—We further analyze the genomewide data of multiple organisms, including *E. coli* [30,86,91], *S. cerevisiae* [29,92], *Schizosaccharomyces pombe* (fission yeast) [31], *Mus musculus* (mouse) [32,93], and *Homo sapiens* (human) [32,94]. Intriguingly, all the measured exponents between m and p are above 0.5 [Figs. 4(a)–4(d) and Fig. S7], 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. Meanwhile, in all the eukaryotic organisms we analyze, their γ 's are still smaller than 1 (Figs. 3 and 4). According to Eq. (8), this indicates that translational regulation also significantly contributes to protein levels by amplifying the effect of differential mRNA levels.

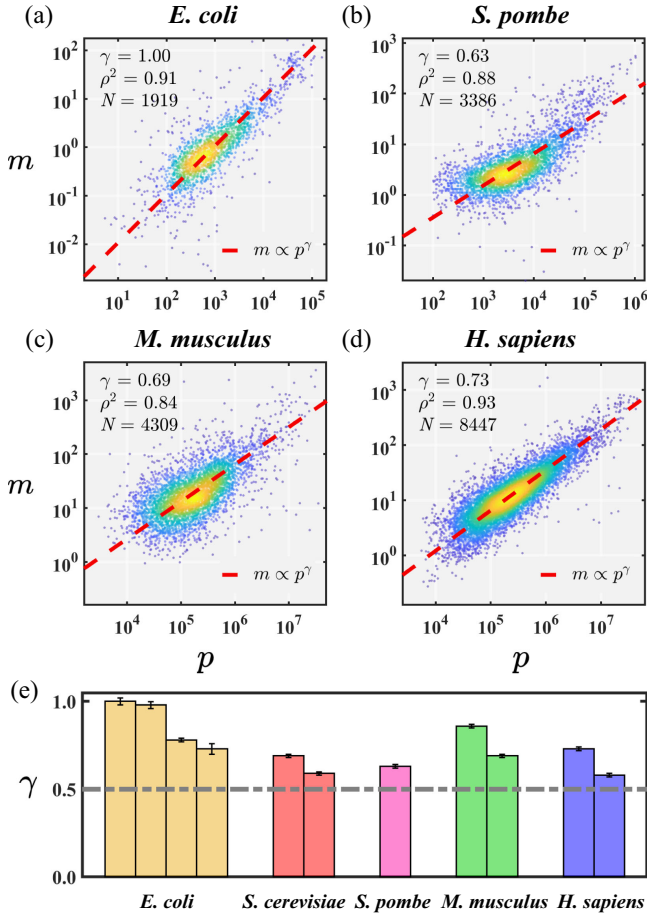


FIG. 4. The m - p scaling across multiple organisms. (a)–(d) Experimental data on m vs p across multiple organisms. Data resources are as follows: *E. coli* [30], *S. pombe* [31], *M. musculus* (NIH 3T3 cell) [93] (the bias in protein measurement corrected by [33]), and *H. sapiens* (HeLa cell) [32]. 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 organisms are all higher than 0.5. Each bar represents one dataset, where the original m - p relations are shown in (a)–(d) and Fig. S7. The error bars represent the standard deviation from bootstrapping (see more statistical details in Sec. D of the Supplemental Material [40]).

We note that the γ exponent can be different for the same organism [Fig. 4(e)]. The different exponents may be due to different strains and conditions across experiments (Table S3), leading to distinct fitness functions and evolutionary outcomes. Other possible sources include different techniques and genomic coverages.

Discussion—We observe two scenarios of gene regulation. In Scenario 1, such as *S. cerevisiae*, the protein toxicity of most proteins is mild, so the scaling exponent is close to—but still higher than—the predicted lower bound, 0.5. In this scenario, transcriptional regulation is primary, but translational regulation still plays a significant

role [29,32,37–39,95,96]. Consistently, Eichhorn *et al.* found that mRNA destabilization is the primary mechanism for miRNA-mediated regulation in multiple eukaryotic cells, and translational repression has a minor but non-negligible contribution [32].

In Scenario 2, notably *E. coli*, the γ exponent is close to 1, and transcriptional regulation appears dominant. According to our theory, the toxicity effects are likely to happen near the expression levels of wild-type cells, which can be tested by systematically measuring fitness as a function of protein level across multiple genes, particularly when the proteins are overexpressed. Regarding the transcriptional dominance in *E. coli*, Balakrishnan *et al.* have suggested that the protein concentration is essentially set by the promoter strength in *E. coli*, given that mRNA degradation rates and translational characteristics are similar across most mRNAs [30], and that mRNA synthesis relies on the ability of promoter to recruit limiting RNA polymerases [77,79,97]. We note that translational regulation may be significant for *E. coli* in some conditions [Figs. S7(b) and S7(c)] [91,98]. Intriguingly, our Letter unifies two scenarios of gene expression regulation from an evolutionary perspective. In all organisms we analyze, the exponents are above the predicted lower bound of 0.5. Our theory does not exclude the possibility of a γ exponent larger than 1.

The method of optimizing fitness with respect to the mRNA level has been applied to determine an upper bound on transcription rates [98]. In Ref. [5], the authors built a mathematical model to explore the trade-off between metabolic load and growth robustness and found that optimal protein expression levels for many genes are vastly overabundant and predicted a linear relationship between the translation efficiency and mRNA copy number, aligning with our model's $m \sim p^{0.5}$ scaling. We remark that our evolutionary minimal model predicts the optimal correlation between m and p , while the microscopic mechanism that generates this correlation can be various, which is not the focus of our minimal model. Essentially, the fitness-driven scaling law does not contradict the constraints imposed by gene regulatory networks. Rather, it reveals the optimized target mRNA and protein levels that govern the evolution of the regulatory networks.

In our Letter, we do not consider the cost of other types of RNA (e.g., miRNA) because they are neither explicitly included nor under optimization in our minimal model. Our minimal model can be adapted to other fitness measures, such as survival ability under stress. Our main conclusions should remain the same for alternative definitions of fitness.

Acknowledgments—The research was funded by the National Key Research and Development Program of China (No. 2024YFA0919600) and supported by Peking-Tsinghua Center for Life Sciences grants.

Data availability—The experimental data analyzed in this study are publicly available in the cited references. The source codes are available upon reasonable request from the authors.

- [1] A. Wagner, Energy constraints on the evolution of gene expression, *Mol. Biol. Evol.* **22**, 1365 (2005).
- [2] E. Dekel and U. Alon, Optimality and evolutionary tuning of the expression level of a protein, *Nature (London)* **436**, 588 (2005).
- [3] M. Kafri, E. Metzl-Raz, G. Jona, and N. Barkai, The cost of protein production, *Cell Rep.* **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, *Sci. Adv.* **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 Syst.* **12**, 924 (2021).
- [7] J. Paulsson, Summing up the noise in gene networks, *Nature (London)* **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 (London)* **441**, 840 (2006).
- [10] A. Sanchez, S. Choubey, and J. Kondev, Regulation of noise in gene expression, *Annu. Rev. Biophys.* **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 Syst.* **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, *Biotechnol. Bioeng.* **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, *J. Bacteriol.* **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, *Can. J. Microbiol.* **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, *Proc. Natl. Acad. Sci. U.S.A.* **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*, *Nat. Commun.* **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. Oplente, J. Fisher, S. Moon, K. Bullaughey, L. B. Carey, and D. Dedousis, Nonlinear fitness consequences of variation in expression level of a eukaryotic gene, *Mol. Biol. Evol.* **30**, 448 (2012).
- [21] F. Duveau, W. Toubiana, and P. J. Wittkopp, Fitness effects of cis-regulatory variants in the *Saccharomyces cerevisiae* TDH3 promoter, *Mol. Biol. Evol.* **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, *Proc. Natl. Acad. Sci. U.S.A.* **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 Syst. Biol.* **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, *Proc. Natl. Acad. Sci. U.S.A.* **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, *Nat. Commun.* **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 Rep.* **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 micromRNAs by the time substantial repression ensues, *Mol. 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, *Nat. Rev. Genet.* **21**, 630 (2020).
- [37] C. Vogel, R. de Sousa Abreu, D. Ko, S.-Y. Le, B. A. Shapiro, S. C. Burns, D. Sandhu, D. R. Boutz, E. M. Marcotte, and L. O. Penalva, Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line, *Mol. Syst. Biol.* **6**, 400 (2010).
- [38] G. Csárdi, A. Franks, D. S. Choi, E. M. Airolidi, and D. A. Drummond, Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast, *PLoS Genet.* **11**, e1005206 (2015).
- [39] B. Ho, A. Baryshnikova, and G. W. Brown, Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome, *Cell Syst.* **6**, 192 (2018).
- [40] See Supplemental Material <http://link.aps.org/supplemental/10.1103/gkzp-r681> for more details on the fitness functions, numerical simulations, and statistics of all datasets used, which also includes Refs. [41–73].
- [41] T. Kalisky, E. Dekel, and U. Alon, Cost–benefit theory and optimal design of gene regulation functions, *Phys. Biol.* **4**, 229 (2007).
- [42] R. Milo and R. Phillips, *Cell Biology by the Numbers* (Garland Science, New York, NY, USA, 2015).
- [43] P. Eser, C. Demel, K. C. Maier, B. Schwalb, N. Pirkel, D. E. Martin, P. Cramer, and A. Tresch, Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression, *Mol. Syst. Biol.* **10**, 717 (2014).
- [44] B. Neymotin, R. Athanasiadou, and D. Gresham, Determination of *in vivo* RNA kinetics using RATE-seq, *RNA* **20**, 1645 (2014).
- [45] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken *et al.*, Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast, *Mol. Syst. Biol.* **7**, 458 (2011).
- [46] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, Global proteome turnover analyses of the yeasts *S. cerevisiae* and *S. pombe*, *Cell Rep.* **9**, 1959 (2014).
- [47] M. Martin-Perez and J. Villén, Determinants and regulation of protein turnover in yeast, *Cell Syst.* **5**, 283 (2017).
- [48] R. Christiano, H. Arlt, S. Kabatnik, N. Mejhert, Z. W. Lai, R. V. Farese, and T. C. Walther, A systematic protein turnover map for decoding protein degradation, *Cell Rep.* **33**, 108378 (2020).
- [49] J. M. Pratt, J. Petty, I. Riba-Garcia, D. H. Robertson, S. J. Gaskell, S. G. Oliver, and R. J. Beynon, Dynamics of protein turnover, a missing dimension in proteomics, *Mol. Cell. Proteomics* **1**, 579 (2002).
- [50] F. Miura, N. Kawaguchi, M. Yoshida, C. Uematsu, K. Kito, Y. Sakaki, and T. Ito, Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs, *BMC Genomics* **9**, 574 (2008).
- [51] H. F. Lodish, A. Berk, C. Kaiser, M. Krieger, A. Bretscher, H. L. Ploegh, K. C. Martin, M. B. Yaffe, and A. Amon, *Molecular Cell Biology* (Macmillan, New York, 2021).
- [52] D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, and F. H. Arnold, Why highly expressed proteins evolve slowly, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14338 (2005).
- [53] M. Oeffinger and D. Zenklusen, To the pore and through the pore: A story of mRNA export kinetics, *Biochim. Biophys. Acta (BBA)-Gene Regul. Mech.* **1819**, 494 (2012).
- [54] U. Raue, S. Oellerer, and S. Rospert, Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence, *J. Biol. Chem.* **282**, 7809 (2007).
- [55] E. M. Phizicky and A. K. Hopper, tRNA biology charges to the front, *Genes Dev.* **24**, 1832 (2010).
- [56] E. Metzl-Raz, M. Kafri, G. Yaakov, I. Soifer, Y. Gurvich, and N. Barkai, Principles of cellular resource allocation revealed by condition-dependent proteome profiling, *eLife* **6**, e28034 (2017).
- [57] T. Borggreffe, R. Davis, A. Bareket-Samish, and R. D. Kornberg, Quantitation of the RNA polymerase II transcription machinery in yeast, *J. Biol. Chem.* **276**, 47150 (2001).
- [58] A. Riba, N. Di Nanni, N. Mittal, E. Arhné, A. Schmidt, and M. Zavolan, Protein synthesis rates and ribosome occupancies reveal determinants of translation elongation rates, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15023 (2019).
- [59] T. von der Haar, A quantitative estimation of the global translational activity in logarithmically growing yeast cells, *BMC Syst. Biol.* **2**, 87 (2008).
- [60] A. M. Edwards, C. M. Kane, R. A. Young, and R. D. Kornberg, Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*, *J. Biol. Chem.* **266**, 71 (1991).
- [61] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, Real-time observation of transcription initiation and elongation on an endogenous yeast gene, *Science* **332**, 475 (2011).
- [62] V. Pelechano, S. Chavez, and J. E. Perez-Ortin, A complete set of nascent transcription rates for yeast genes, *PLoS One* **5**, e15442 (2010).
- [63] B. Futcher, G. Latter, P. Monardo, C. McLaughlin, and J. Garrels, A sampling of the yeast proteome, *Mol. Cell. Biol.* **19**, 7357 (1999).
- [64] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg, and S. Linnarsson, Quantitative single-cell RNA-seq with unique molecular identifiers, *Nat. Methods* **11**, 163 (2014).

- [65] I. Jarmoskaite, S. K. Denny, P. P. Vaidyanathan, W. R. Becker, J. O. Andreasson, C. J. Layton, K. Kappel, V. Shivashankar, R. Sreenivasan, R. Das *et al.*, A quantitative and predictive model for RNA binding by human pumilio proteins, *Mol. Cell* **74**, 966 (2019).
- [66] G. V. Dolgalev, T. A. Safonov, V. A. Arzumanyan, O. I. Kiseleva, and E. V. Poverennaya, Estimating total quantitative protein content in *escherichia coli*, *saccharomyces cerevisiae*, and *hela* cells, *Int. J. Mol. Sci.* **24**, 2081 (2023).
- [67] J. W. Posakony, J. M. England, and G. Attardi, Mitochondrial growth and division during the cell cycle in *hela* cells, *J. Cell Biol.* **74**, 468 (1977).
- [68] S. B. Cambridge, F. Gnad, C. Nguyen, J. L. Bermejo, M. Kruger, and M. Mann, Systems-wide proteomic analysis in mammalian cells reveals conserved, functional protein turnover, *J. Proteome Res.* **10**, 5275 (2011).
- [69] 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).
- [70] R. C. Jansen, J.-P. Nap, and L. Mlynárová, Errors in genomics and proteomics, *Nat. Biotechnol.* **20**, 19 (2002).
- [71] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, *SIAM J. Numer. Anal.* **17**, 883 (1980).
- [72] I. Markovsky and S. Van Huffel, Overview of total least-squares methods, *Signal Processing* **87**, 2283 (2007).
- [73] N. Draper, *Applied Regression Analysis* (McGraw-Hill, Inc, New York, 1998).
- [74] P. Shah, Y. Ding, M. Niemczyk, G. Kudla, and J. B. Plotkin, Rate-limiting steps in yeast protein translation, *Cell* **153**, 1589 (2013).
- [75] F. Ceroni, R. Algar, G.-B. Stan, and T. Ellis, Quantifying cellular capacity identifies gene expression designs with reduced burden, *Nat. Methods* **12**, 415 (2015).
- [76] 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 Biotechnol.* **34**, 652 (2016).
- [77] J. Lin and A. Amir, Homeostasis of protein and mRNA concentrations in growing cells, *Nat. Commun.* **9**, 4496 (2018).
- [78] Y. Han and F. Zhang, Control strategies to manage trade-offs during microbial production, *Curr. Opin. Biotechnol.* **66**, 158 (2020).
- [79] Q. Wang and J. Lin, Heterogeneous recruitment abilities to RNA polymerases generate nonlinear scaling of gene expression with cell volume, *Nat. Commun.* **12**, 6852 (2021).
- [80] 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-Lamothé *et al.*, RNA polymerase II dynamics and mRNA stability feedback scale mRNA amounts with cell size, *Cell* **186**, 5254 (2023).
- [81] L. Calabrese, L. Ciandrini, and M. Cosentino Lagomarsino, How total mRNA influences cell growth, *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2400679121 (2024).
- [82] 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).
- [83] 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, *Mol. Cell* **65**, 142 (2017).
- [84] Y. Yan and J. Lin, Disentangling the fitness cost of gene expression, *bioRxiv* (2025).
- [85] J. Paulsson, Models of stochastic gene expression, *Phys. Life Rev.* **2**, 157 (2005).
- [86] 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).
- [87] 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, *Nat. Genet.* **38**, 636 (2006).
- [88] V. Shahrezaei and P. S. Swain, Analytical distributions for stochastic gene expression, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17256 (2008).
- [89] M. Mundt, A. Anders, S. M. Murray, and V. Sourjik, A system for gene expression noise control in yeast, *ACS Synth. Biol.* **7**, 2618 (2018).
- [90] 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).
- [91] 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).
- [92] F. W. Albert, D. Muzzey, J. S. Weissman, and L. Kruglyak, Genetic influences on translation in yeast, *PLoS Genet.* **10**, e1004692 (2014).
- [93] 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 (London)* **473**, 337 (2011).
- [94] 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, *Mol. Syst. Biol.* **7**, 548 (2011).
- [95] F. Edfors, F. Danielsson, B. M. Hallström, L. Käll, E. Lundberg, F. Pontén, B. Forsström, and M. Uhlén, Gene-specific correlation of RNA and protein levels in human cells and tissues, *Mol. Syst. Biol.* **12**, 883 (2016).
- [96] A. Franks, E. Airolidi, and N. Slavov, Post-transcriptional regulation across human tissues, *PLoS Comput. Biol.* **13**, e1005535 (2017).
- [97] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Transcriptional regulation by the numbers: Models, *Curr. Opin. Genet. Dev.* **15**, 116 (2005).
- [98] J. Hausser, A. Mayo, L. Keren, and U. Alon, Central dogma rates and the trade-off between precision and economy in gene expression, *Nat. Commun.* **10**, 68 (2019).

End Matter

Analytical derivation—We first find the optimal p that maximizes the total fitness before considering the noise effects, where

$$F = \left[1 - b \frac{p_b^{s_b}}{p^{s_b} + p_b^{s_b}} \right] \left[1 - c_m m - c_p p - \left(\frac{p}{p_d} \right)^{s_d} \right] \approx 1 - b \left(\frac{p}{p_b} \right)^{-s_b} - c_m m - c_p p - \left(\frac{p}{p_d} \right)^{s_d}. \quad (\text{A1})$$

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 level p_b [5]. We compute $\partial F(m, p)/\partial p = 0$ and find that

$$b s_b \left(\frac{p_b}{p} \right)^{s_b} - s_d \left(\frac{p}{p_d} \right)^{s_d} - c_p p = 0. \quad (\text{A2})$$

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

$$\frac{\partial^2 F}{\partial p^2} = -\frac{c_p(s_b + 1)}{p} \left[1 + \frac{(s_b + s_d)s_d}{(s_b + 1)c_p p} \left(\frac{p}{p_d} \right)^{s_d} \right]. \quad (\text{A3})$$

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

$$\langle F(m, p) \rangle = F(m, p) + \Delta F. \quad (\text{A4})$$

Combining the above equation with Eqs. (4), (6), and (A3), we find that

$$m = \sqrt{\frac{\tau_m |\partial^2 F / \partial p^2|}{2\tau_p c_m}} p = \sqrt{\alpha \left[1 + \left(\frac{p}{p_c} \right)^{s_d-1} \right]} p^{\frac{1}{2}}. \quad (\text{A5})$$

Here, $\alpha = [c_p \tau_m (s_b + 1) / 2c_m \tau_p]$, and $p_c = \{[(s_b + 1)c_p / (s_b + s_d)s_d] p_d^{s_d}\}^{\frac{1}{s_d-1}}$. The agreement between the theoretical predictions and the numerical calculations of the full model supports the validity of the above approximations.