Universality of phenotypic distributions in bacteria

Kuheli Biswas and Naama Brenner
Dept. of Chemical Engineering and Network Biology Research lab,
Technion - Israel Institute of Technology, Haifa, Israel

Some phenotypic properties in bacteria exhibit universal statistics, with distributions collapsing under scaling. The extent and origins of such universality are not well understood. Using phenomenological modeling of growth and division, we identify compound "shape-factors" that describe the distributions throughout a large set of single-cell data. We find that the emergence of universal distributions is associated with the robustness of shape-factors across conditions, explaining the universality of cell size and highly expressed protein content and the non-universality of times between consecutive divisions. A wide range of experimental data sets support our theory quantitatively.

Bacterial growth and division have fascinated scientists and have been studied for decades. Mathematical models to account for the variability in cell size and division times across time and among individuals were developed early on [1–4], but the comparison to data at the time was limited. More recent single-cell imaging and microfluidics technology yields high-quality data with large statistical power [5–12], enabling to test quantitative models of growth and division and the relation between temporal dynamics and statistical properties.

Previous works have reported that the distributions of cell size and highly expressed proteins are non-Gaussian skewed, well approximated by a log-normal form [7, 13–16]. These distributions have a universal shape, as manifested by their collapse under scaling [17, 18]. In physics, such distribution collapse is associated with the insensitivity of macroscopic properties to underlying microscopic ones [19]. In biological systems, analogous collapse appears in different contexts, but its origin and extent are still not well understood [20**?** , 21]

The distribution of times between divisions has also been extensively studied, but here the emerging picture is less conclusive. Some experiments report a Gaussian-like distribution of division times [8, 9, 18, 22] while others find more skewed distributions [1, 2, 12, 23–25]. Some of these report a universal shape for division times [26, 27]. Modeling work assuming a small-noise approximation has concluded that times between divisions are distributed approximately Gaussian [15]. In contrast, "sloppy size control" models [28] as well as other modeling approaches [29–31] find skewed distributions. Thus, a unifying framework that can explain the observed diversity and apparent disagreement in division time distribution shapes is lacking.

Here, we compute the division time distribution using *first passage time* techniques in a phenomenological coarse-grained stochastic model of growth and division. We find that the ratio between two stochastic parameters - variability in exponential growth rate and in threshold - defines a dimensionless "shape-factor" that governs the distribution shape beyond shift and scale transformations. Analyzing a large set of single-cell data from microfluidic experiments, we find that division time distributions span a range of diverse behaviors, from almost Gaussian to very skewed. Our theoretical predictions, based on direct estimation of dynamic parameters, agree with the data throughout this range of behaviors. Combining these observations with known properties of cell size and highly expressed protein content, we address the question of universality. By observing the spread of data points in the landscape of shape-factors, we find that for cell size and protein, the data adhere near a manifold of fixed value, while for division times they scatter through multiple contour lines. This finding sheds light on the universality of size and protein distributions, and the non-universality (or shape diversity) of division time distributions. It highlights the emergence of compound parameter combinations, whose sensitivity is key to understanding the system's behavior.

Model Definition. High-resolution single-cell measurements have revealed that several bacterial species, including *E. coli.* grow smoothly and exponentially across each cycle [8, 24, 25]; division occurs abruptly - consistent with the idea of triggering by some threshold event (Fig. 1 a). The growth dynamics during the nth cycle can thus be approximated as

$$x_n(t) = x_n^b\, e^{\alpha_n t} = x_n^b\, e^{\phi_n(t)}, \quad 0 < t < T_n. \tag{1}$$

Here x_n is the cell size or highly expressed protein content, with x_n^b the value at birth. α_n is the effective growth rate, and $\phi_n(t)$ the Logarithmic Fold Change (LFC) in cell size. The cycle ends at time $t = T_n$ with symmetric division. We formulate a generalized stochastic threshold-crossing model of division that unifies different previously studied division mechanisms as special cases.

The LFC along the nth cycle is by definition zero at birth $\phi_n(0) = 0$, and at division reaches a fluctuating values ϕ_{c_n} (Fig. 1a). These fluctuations are well described by a Gaussian distribution around $\ln 2$ (Fig. 1b), with negligible correlations across cycles (Fig. 1c) and a significant negative correlation with cell size at birth (Fig. 1d). These statistical properties are all captured by the following phenomenological expression [24]

$$\phi_{c_n} = \ln 2 - \beta \ln\left(\tilde{x}_n^b\right) + \xi_n. \tag{2}$$

Here ξ_n is a Gaussian white noise with zero mean, $\tilde{x}_n^b$ is the birth size normalized by its mean, and $0 < \beta \leq 1$ is

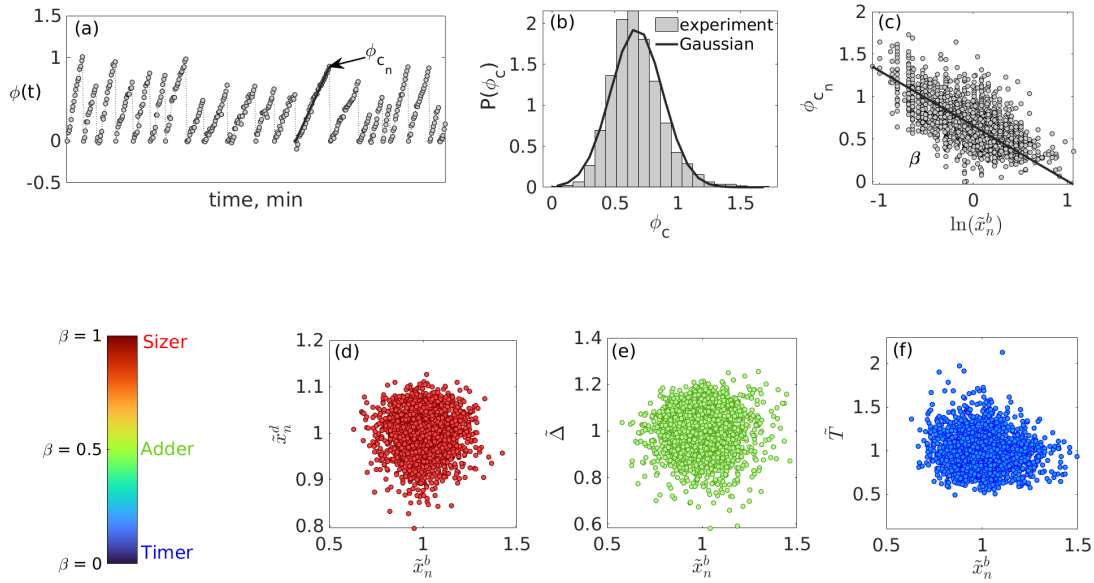

FIG. 1. **Phenomenological model unifying different division control mechanism. (a-d)**: Data from [25], where *E. coli* were grown in a microfluidic device. **(a)** In exponential growth, Logarithmic Fold Change (LFC) in cell size, $\phi(t) = \alpha t$, is linear across cycles. **(b)** At division, LFC reaches a stochastic threshold ϕ_c, with distribution $P(\phi_c)$ approximately Gaussian ($\mu_\phi \approx \ln 2$, $\sigma_\phi = 0.17$). **(c)** Correlations in ϕ_c across consecutive cycles are negligible with correlation coefficient $\approx$ -0.02. **(d)** The threshold ϕ_c is negatively correlated with normalized birth size $\tilde{x}_b$ with best-fit slope $\beta = 0.65$. Experimental details are provided in [32]. **(e-g)** Scatterplots from simulations of the phenomenological model, Eqs. 1, 2: division size (d), added size (e) and cycle time (f) vs. normalized initial size. The independence of these variables on initial condition is often interpreted as indicating a control mechanism where they are kept constant. By tuning the homeostasis parameter β, the model reduces to previously studied modes of division control: **(d)** sizer ($\beta = 0.9$), **(e)** adder ($\beta = 0.5$), and **(f)** timer ($\beta = 0.2$) as special cases. 10^4 generations were simulated with $\alpha = 0.023$ min^{-1} and $\xi = \mathcal{N}(0, 0.15)$. See [32] for more details.

the slope of Fig. 1d. Due to the lognormal distribution of size, the correlation term does not modify the Gaussian nature of ϕ_c; at the same time, such a negative correlation induces a stable size distribution over multiple generations, interpolating between previously described division models [15, 16, 24, 33, 34]. To demonstrate this, we simulate the model Eqs. (1, 2) and present the resulting scatterplots of different variables vs. initial cell size in Fig. 1(e-g). Such plots have been extensively used to identify the mode of division control. For example in Fig. 1f, $\beta = 0.5$ and added size is independent of the birth size, a correlation known as the "adder" mode of division control. Sizer and timer-like correlations are found for other values of β in Fig. 1(e,g). The mathematical mapping between the current and previously formulated models is elaborated in [32].

Model Solution. Eqs. (1, 2) together constitute a phenomenological model that was successfully solved for cell size and protein distributions [15, 16], and will be analyzed below in terms of division time statistics. We assume that cell division occurs when the LFC crosses a fluctuating process; this does not necessarily imply a mechanistic interpretation, but rather a tool for calculation. Using the first passage time (FPT) technique [35, 36]. Cell division occurs when the Gaussian process $P(\phi_c) = \mathcal{N}(\ln 2, \sigma_\phi)$ crosses the absorbing boundary $\phi(t) = \alpha t$ [37]. The survival probability - the probability that the cell has not divided till time t - can then be expressed as

$$S(t) = \int_{\phi(t)}^{\infty} P(\phi_c) d\phi_c = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\phi(t) - \ln 2}{\sqrt{2\sigma_\phi^2}}\right)\right]. \tag{3}$$

For a fixed growth rate α, this is related to the probability density of the FPT by

$$P(T|\alpha) = \frac{\alpha}{\sqrt{2\pi\sigma_\phi^2}} exp\left[-\frac{(\alpha T - \ln 2)^2}{2\sigma_\phi^2}\right]. \tag{4}$$

However, the growth rate estimated from the slope of $\phi(t)$ varies across cycles and follows Gaussian distribution, $P(\alpha) = \mathcal{N}(\mu_\alpha, \sigma_\alpha)$ (Fig. 2a). To account for variable growth rates, we integrate Eq. (4) with $P(\alpha)$, which leads to the following division time distribution [32]

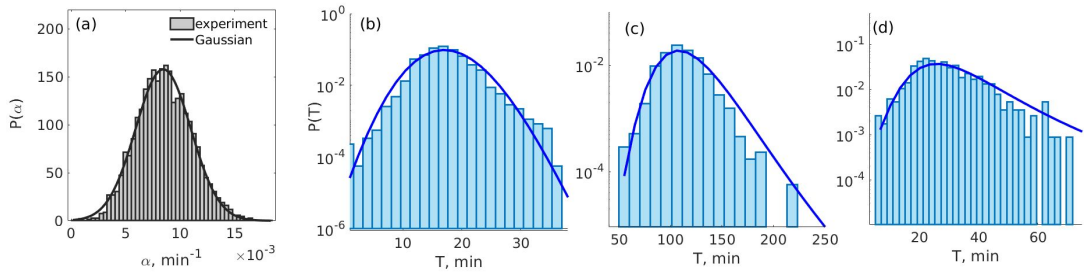

FIG. 2. **Division time distributions are predicted across a range of shape-factors. (a)** Growth rates α_n, obtained from the slope of $\phi_n(t)$ (Fig. 1a) follow a Gaussian distribution, $P(\alpha) = \mathcal{N}(\mu_\alpha, \sigma_\alpha)$. These data from [25] show $\chi_\alpha = \sigma_\alpha/\mu_\alpha = 0.27$. **(b)-(d)** Division time distributions from data (bars) corresponding to different Σ_T are well predicted by Eq. (5) (solid lines). The shape factors are (b) $\Sigma_T = 0.64$ [8], glucose medium, (c) $\Sigma_T = 1.3$ [11] arginine medium, (d) $\Sigma_T = 1.8$ [24], LAC-M9 medium. In all experiments, *E. coli* bacteria were grown in mother machines; see details in [32].

$$P(T) = \int P(T|\alpha)P(\alpha)d\alpha = \frac{(\mu_\alpha\sigma_\phi^2 + \sigma_\alpha^2 T\ln 2)}{\sqrt{2\pi}(\sigma_\phi^2 + (\sigma_\alpha T)^2)^{3/2}}\ \exp\left[-\frac{(\mu_\alpha T - \ln 2)^2}{2(\sigma_\phi^2 + (T\sigma_\alpha)^2)}\right]. \tag{5}$$

Transforming this expression to dimensionless time [32], we find that two dimensionless parameters appear in the formula for division times: σ_ϕ, a measure of threshold variance, and $\chi_\alpha = \sigma_\alpha/\mu_\alpha$, a measure of growth-rate variability. The shape of the distribution is strongly influenced by the ratio of these two parameters, which reflects the relative strength of the two noise sources in the model: $\Sigma_T = \chi_\alpha/\sigma_\phi$. We term this ratio the shape-factor; it defines regions where the distribution resembles a Gaussian shape when $\Sigma_T < 1$, and a skewed heavy-tailed shape when $\Sigma_T > 1$.

Fig. 2(b-d) shows a few examples taken from different experiments, illustrating a range of shape factors from $\Sigma_T = 0.64$, an almost Gaussian shape, to a larger value of $\Sigma_T = 1.8$, where the shape is highly skewed. The experimental data (bars) are well described by Eq. 5 (lines) throughout this range of behaviors. A larger collection of division time distributions can be seen in [32], all in good agreement with the theory. Note that there are no fitting parameters in this theory. To obtain the theoretical lines, we estimate stochastic variables μ_α, σ_α and σ_ϕ directly from dynamic trajectories such as those in Fig. 1(a). These are then inserted into Eq. (5).

Universal and non-universal distributions. The stochastic LFC threshold model was used in previous work to predict a log-normal distribution of cell size and protein content; its shape depends on the compound shape-factor $\Sigma_X = \sigma_\phi/\sqrt{\beta(2-\beta)}$, which in turn depends on the threshold standard deviation σ_ϕ and the homeostasis parameter β [15, 16]. Here we identified the compound parameter $\Sigma_T = \sigma_\phi/\chi_\alpha$ as determining the shape of the division time distribution. In contrast to the broad range of distribution shapes found for division time, cell size and highly expressed protein content have been consistently reported to exhibit a universal distribution shape[14, 15, 17, 18, 27]. We next investigate properties of the shape-factors to shed light on the question of universal distribution collapse.

In Fig. 3(a,b) we assemble data from a large set of single-cell experiments in the planes of stochastic parameters $(\chi_\alpha, \sigma_\phi)$ and (β, σ_ϕ) respectively. Each point represents one experiment, where single-cell traces were pooled for estimation. In all cases, bacteria were grown in mother machines, but relevant conditions such as medium, temperature, or bacterial strain vary significantly among experiments (see Table 1 in [32]). In addition, in these parameter planes, the shape-factor for division time (a) and cell size (b) are depicted as grey contour lines of constant values. Examining the embedding of the data across the shape landscape reveals an important difference between the cases.

In (a), the data scatter across many contour lines, spanning a broad range of shape-factor values ($\Sigma_T \sim 0.1$ to 3). Contributing to this variability is the broad range of growth rate noise χ_α (x-axis), approximately ten-fold 0.05 to 0.5, to which the shape-factor is inversely proportional. This is consistent with the broad range of behaviors we observed for division time distributions across data sets. Two examples of scaled distributions are shown in panel (c). In (b), in marked contrast, the data are concentrated in a small region around a single contour line, maintaining an almost constant shape-factor for cell size ($\Sigma_X \approx 0.2$). It has previously been demonstrated that the shape of the lognormal distribution is very weakly sensitive to the shape-parameter value in this regime [16]; together these observations are consistent with the collapse of these distributions upon scaling (panel (d)).

Discussion. We examined single-cell statistical properties using a phenomenological growth and division model. This model relies on several key assumptions: exponential growth during the cell cycle, variable exponential

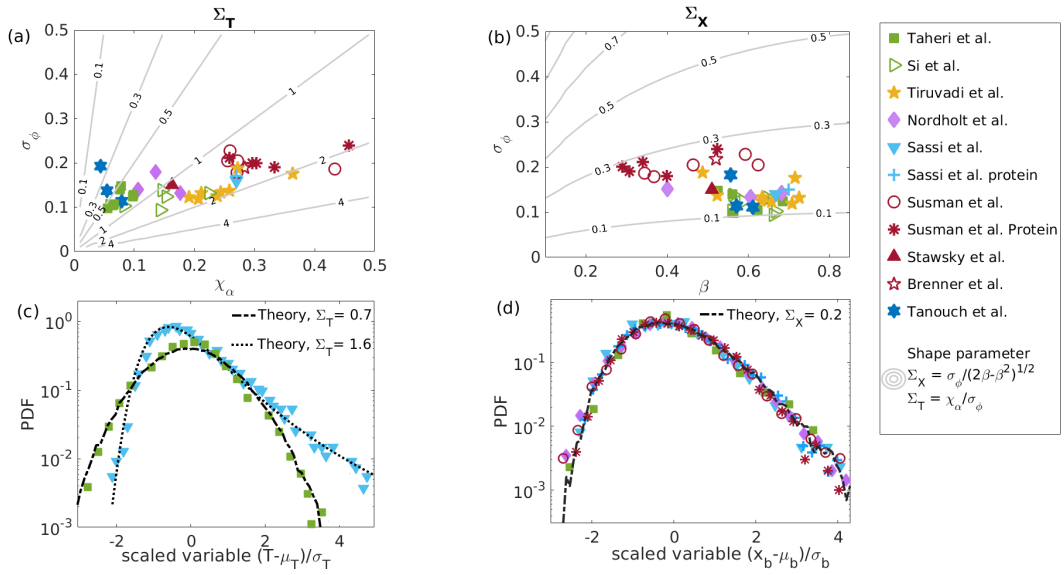

FIG. 3. **Non-universal division time distribution and universal cell size distribution.** **(a)** Plot of χ_α vs σ_ϕ, estimated from experimental data, span a broad range of shape factors Σ_T (gray contours). **(b)** Plot of σ_ϕ vs. β, estimated for the same data, cluster around a small range of shape factor for cell size or protein Σ_X (gray contours). **(c)** The scaled distribution of cell division times is non-universal, showing a Gaussian-like distribution for $\Sigma_T < 1$ (green squares, [8]) and a skewed shape for $\Sigma_T > 1$ (blue triangles, [25]). The black curves are obtained from our theory Eq. (5). **(d)** The scaled distribution of cell size and protein number collapse under two-parameter scaling ([8, 22, 24, 25]), and agree well with log-normal distribution [16] (black curve). Data are taken from [7–12, 22, 24, 25, 38] and experimental details are provided in Table 1 of [32].

rates in successive cycles, and symmetric division triggered by a noisy threshold in Log Fold Change of cell size - negatively correlated with birth size. This approach has been validated in previous studies, and encompasses various homeostasis modes, including adder, sizer, and timer scenarios, as specific instances. [39]

By employing a first passage time framework, we computed the distribution of division times. The approximations made in obtaining this result do not hinge on a small-noise approximation in growth rate across cell cycles (see [32]). In the data we analyzed, encompassing a wide spectrum of experiments in microfluidic traps, the small-noise limit only applies to a minority of experiments. The range of behaviors - from very narrow to very broad growth rate distribution - served as a central motivation for extending the phenomenological theory in our present study.

The resulting distribution has strictly no finite moments, but a limited range of data allows for a good comparison with an appropriate cutoff. Notably, this result accurately predicts the relationships between data moments and model parameters, as detailed in [32]. The long tails of the distribution result from the Gaussian assumption for growth rates. Alternative assumptions, such as a Gamma distribution, yield empirically similar distributions [32], but do not provide an analytic expression that allows for identifying parameter dependencies and a clearly understandable shape-factor.

The shape-factor Σ_T governs the distribution shape beyond shift and scale, eliminated by two-parameter scaling. An analogous factor had previously been found for cell size, but its empirical values and significance for distribution scaling were not studied. By comparing the two shape-factors we were able to illuminate the sources of distribution universality, as stemming from the robustness of these parameter combinations as experimental conditions are varied. This explanation remains phenomenological and does not clarify the source of difference in robustness. In particular, the reasons behind different levels of variability in single-cell growth rates among experimental conditions remain to be understood. It can be speculated that the sensitivity of growth rates to their environment and their strong correlation with division times over long timescales [38] make them effective control variables that require flexibility for compensation. On the other hand, cell size does not partake in long-term homeostatic correlations; stable growth can occur at various cell sizes, rendering it an "irrelevant variable" in the sense of statistical mechanics with a universal distribution. These hypotheses warrant further investigation.

The description of cell size and highly expressed proteins has converged to a similar modeling framework. Both quantities adhere to the fundamental model assumptions outlined above and exhibit strong cycle-by-cycle correlations between their stochastic variables [7, 24, 25]. Consequently, in Fig. 3 we found that they share the same universal distribution shape. Intriguingly, the statistics of division time can be predicted based on the dynamics of either cell size or any highly expressed protein, with similar predictive success [32]. Thus, while cell

division is likely triggered by multiple events, their tight interconnection reduces the complexity of the problem, allowing several options to predict statistical properties effectively. This aligns with previous observations of balanced exponential growth as a dynamic attractor characterized by robust relationships among constituents [31]. Understanding such reduction in effective dimensionality in high-dimensional yet highly correlated systems like the biological cell remains an important open question.

Acknowledgements: K.B. is supported by a fellowship from the Lady Davis Foundation at the Technion. The research is supported in part by the Israeli Science Foundation (grant number 155/18), and the Binational Science Foundation (grant number 2016376). We are grateful to Erez Braun and Ron Teichner for comments on the manuscript.

AI declaration: While preparing this work, the authors used ChatGPT to improve the readability and language of the work. After using this tool, we reviewed and edited the content as needed and take full responsibility for the paper's content.

[1] A. Koch and M. Schaechter, A model for statistics of the cell division process, Microbiology **29**, 435 (1962).
[2] E. Powell and F. Errington, Generation times of individual bacteria: some corroborative measurements, Microbiology **31**, 315 (1963).
[3] A. G. Marr, R. Harvey, and W. Trentini, Growth and division of Escherichia coli, Journal of Bacteriology **91**, 2388 (1966).
[4] W. Voorn and L. Koppes, Skew or third moment of bacterial generation times, Archives of microbiology **169**, 43 (1997).
[5] F. Wu and C. Dekker, Nanofabricated structures and microfluidic devices for bacteria: from techniques to biology, Chemical Society Reviews **45**, 268 (2016).
[6] P. Wang, L. Robert, J. Pelletier, W. L. Dang, F. Taddei, A. Wright, and S. Jun, Robust growth of Escherichia coli, Current biology **20**, 1099 (2010).
[7] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman, Single-cell protein dynamics reproduce universal fluctuations in cell populations, The European Physical Journal E **38**, 1 (2015).
[8] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun, Cell-size control and homeostasis in bacteria, Current biology **25**, 385 (2015).
[9] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You, A noisy linear map underlies oscillations in cell size and gene expression in bacteria, Nature **523**, 357 (2015).
[10] D. Yang, A. D. Jennings, E. Borrego, S. T. Retterer, and J. Männik, Analysis of factors limiting bacterial growth in PDMS mother machine devices, Frontiers in microbiology **9**, 871 (2018).
[11] F. Si, G. Le Treut, J. T. Sauls, S. Vadia, P. A. Levin, and S. Jun, Mechanistic origin of cell-size control and homeostasis in bacteria, Current Biology **29**, 1760 (2019).
[12] S. Tiruvadi-Krishnan, J. Männik, P. Kar, J. Lin, A. Amir, and J. Männik, Coupling between DNA replication, segregation, and the onset of constriction in Escherichia coli, Cell reports **38**, 110539 (2022).
[13] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko, Ubiquity of log-normal distributions in intracellular reaction dynamics, Biophysics **1**, 25 (2005).
[14] K. Hosoda, T. Matsuura, H. Suzuki, and T. Yomo, Origin of lognormal-like distributions with a common width in a growth and division process, Physical Review E **83**, 031118 (2011).
[15] A. Amir, Cell size regulation in bacteria, Physical review letters **112**, 208102 (2014).
[16] N. Brenner, C. Newman, D. Osmanović, Y. Rabin, H. Salman, and D. Stein, Universal protein distributions in a model of cell growth and division, Physical Review E **92**, 042713 (2015).
[17] H. Salman, N. Brenner, C. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, Universal protein fluctuations in populations of microorganisms, Physical review letters **108**, 238105 (2012).
[18] S. Iyer-Biswas, C. S. Wright, J. T. Henry, K. Lo, S. Burov, Y. Lin, G. E. Crooks, S. Crosson, A. R. Dinner, and N. F. Scherer, Scaling laws governing stochastic growth and division of single bacterial cells, Proceedings of the National Academy of Sciences **111**, 15912 (2014).
[19] A. Cohen, Y. Roth, and B. Shapiro, Universal distributions and scaling in disordered systems, Physical Review B **38**, 12125 (1988).
[20] L.-h. So, A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, General properties of transcriptional time series in escherichia coli, Nature genetics **43**, 554 (2011).
[21] H.-Y. Shih, H. Mickalide, D. T. Fraebel, N. Goldenfeld, and S. Kuehn, Biophysical constraints determine the selection of phenotypic fluctuations during directed evolution, Physical biology **15**, 065003 (2018).
[22] N. Nordholt, J. H. van Heerden, and F. J. Bruggeman, Biphasic cell-size and growth-rate homeostasis by single bacillus subtilis cells, Current Biology **30**, 2238 (2020).
[23] J. L. Lebowitz and S. Rubinow, A theory for the age and generation time distribution of a microbial population, Journal of Mathematical Biology **1**, 17 (1974).
[24] L. Susman, M. Kohram, H. Vashistha, J. T. Nechleba, H. Salman, and N. Brenner, Individuality and slow dynamics in bacterial growth homeostasis, Proceedings of the National Academy of Sciences **115**, E5679 (2018).
[25] A. S. Sassi, M. Garcia-Alcala, M. Aldana, and Y. Tu, Protein Concentration Fluctuations in the High Expression Regime: Taylor's Law and Its Mechanistic Origin, Physical review X **12**, 011051 (2022).
[26] S. Iyer-Biswas, F. Hayot, and C. Jayaprakash, Stochasticity of gene products from transcriptional pulsing, Physical Review E **79**, 031911 (2009).
[27] A. S. Kennard, M. Osella, A. Javer, J. Grilli, P. Nghe, S. J. Tans, P. Cicuta, and M. C. Lagomarsino, Individuality and universality in the growth-division laws of single

e. coli cells, Physical Review E **93**, 012408 (2016).
[28] J. J. Tyson and O. Diekmann, Sloppy size control of the cell division cycle, Journal of theoretical biology **118**, 405 (1986).
[29] H. Kubitschek, The distribution of cell generation times, Cell proliferation **4**, 113 (1971).
[30] R. Pugatch, Greedy scheduling of cellular self-replication leads to optimal doubling times with a log-frechet distribution, Proceedings of the National Academy of Sciences **112**, 2611 (2015).
[31] P. P. Pandey, H. Singh, and S. Jain, Exponential trajectories, cell size fluctuations, and the adder property in bacteria follow from simple chemical dynamics and division control, Physical Review E **101**, 062406 (2020).
[32] See supplemental material at, .
[33] D. A. Kessler and S. Burov, Effective potential for cellular size control, arXiv preprint arXiv:1701.01725 (2017).
[34] P.-Y. Ho, J. Lin, and A. Amir, Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects, Annual review of biophysics **47**, 251 (2018).
[35] C. W. Gardiner, *Handbook of stochastic methods*, Vol. 3 (springer Berlin, 1985).
[36] H. Risken, Fokker-planck equation, in *The Fokker-Planck Equation* (Springer, 1996) pp. 63–95.
[37] L. Luo, Y. Bai, and X. Fu, Stochastic threshold in cell size control, Physical Review Research **5**, 013173 (2023).
[38] A. Stawsky, H. Vashistha, H. Salman, and N. Brenner, Multiple timescales in bacterial growth homeostasis, Iscience **25**, 103678 (2022).
[39] C. Jia, A. Singh, and R. Grima, Cell size distribution of lineage data: analytic results and parameter inference, Iscience **24** (2021).