Annual Review of Biophysics

Modeling Cell Size Regulation: From Single-Cell-Level Statistics to Molecular Mechanisms and Population-Level Effects

Po-Yi Ho, Jie Lin, and Ariel Amir

John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA; email: arielamir@seas.harvard.edu

Annu. Rev. Biophys. 2018. 47:12.1–12.21

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

https://doi.org/10.1146/annurev-biophys-070317-032955

Keywords

cell size regulation, coarse-grained stochastic models, microbial cell cycle, molecular mechanisms, single-cell variability, population growth

Abstract

Most microorganisms regulate their cell size. In this article, we review some of the mathematical formulations of the problem of cell size regulation. We focus on coarse-grained stochastic models and the statistics that they generate. We review the biologically relevant insights obtained from these models. We then describe cell cycle regulation and its molecular implementations, protein number regulation, and population growth, all in relation to size regulation. Finally, we discuss several future directions for developing understanding beyond phenomenological models of cell size regulation.

Contents

1. INTRODUCTION

Most microorganisms regulate their cell size, as evidenced by their narrow cell size distributions. In particular, all known species of bacteria have cell size distributions with a small coefficient of variation (CV) (standard deviation divided by the mean), which can be as low as 0.1 (29). For cells that grow exponentially, a small CV for size implies a small CV for interdivision times. However, a small CV for interdivision times is not sufficient to regulate cell size; as we demonstrate below, a simple timer strategy cannot regulate cell size in the face of fluctuations. Cells must therefore have a way to effectively measure size.

The physiological implications of cell size remain under debate. In the context of bacteria, these implications are discussed in detail in a recent, excellent review (60), which also stresses the intimate connection between the problem of cell size regulation and that of cell cycle regulation. For instance, cell division, which mechanistically determines cell size, is coupled to DNA replication. In this review, we do not focus on the rich biology behind this problem, but instead elaborate on the various phenomenological models developed to study this problem over the past several decades. These are typically coarse-grained models, which consider the cell as a whole and describe cell volume at various stages of the cell cycle. They often seek to capture the statistics of the random process underlying cell size regulation. For example, what is the relationship between cell size and interdivision time, and what distributions characterize the fluctuations in these variables?

A devil's advocate or a biologist may ask why one would care about these questions. Quantitatively describing the distributions and finding scaling relationships between variables are worthy goals from a physicist's statistical mechanical point of view, but can such phenomenological modeling shed light on the biology? Three distinct examples indicate that the answer to this question is affirmative.

First, for several bacterial model species, including *Escherichia coli* and *Bacillus subtilis*, cell volume scales exponentially with growth rate and proportionally with, loosely speaking, chromosome copy number (23, 45, 51). The scaling constant is in fact equal to the time from the initiation of DNA replication to cell division. It was shown 50 years ago that this observation can be rationalized within a model in which the regulation of cell size does not occur via controlling the timing of cell divisions, but rather via controlling the timing of the initiation of DNA replication (13).

In this way, a quantitative pattern on the phenomenological level, with the aid of mathematical modeling, led to an important insight regarding bacterial physiology. The same empirical observation helped to address whether cell size regulation occurs over cell volume, surface area, or other dimensions. Experiments in rod-shaped bacteria often measure cell length and therefore cannot distinguish between these possibilities since cell width in these bacteria is very narrowly distributed (CV < 0.05) (51). As a result, in addition to cell volume, both cell surface area and cell length have been proposed to set cell size (11, 21). However, recent experiments in *E. coli* showed that the same scaling relation holds, but only for cell volume and not for surface area or width, under genetic perturbations to cell dimensions (62). This result supports the idea that volume is the key phenomenological variable controlling cell size. Below, we use the term cell size for generality while keeping the above discussion in mind.

Second, a naive proposal for cell size regulation is a timer strategy, in which cells control the timing of their cell cycles so that, on average, cell size doubles from birth to division. However, it can be shown by theoretical arguments alone that this mode of regulation is incompatible with the small CVs of cell size distributions if cell volume grows exponentially in time at the single-cell level, as seen in experiments (19). This is because the cumulative effect of noise will cause the variance in cell size to diverge. Explicitly, consider exponentially growing cells with a constant growth rate λ and stochastic interdivision time t_d. A cell born at size v_b will generate a progeny of size $v_b' = v_b e^{\lambda t_d}/2$, assuming perfect symmetric division. If we let $x = \ln(v_b/v_0)$ be the log size, where v_0 is a constant that sets the mean cell size, and x' be the log size at the next generation, then

$$x' = x + \lambda t_d - \ln 2. \qquad 1.$$

Uncorrelated fluctuations in t_d will then lead to a random walk in log sizes with fluctuations accumulating as the square root of the number of divisions. Thus, the cell size distribution in a growing population will not reach stationarity via a timer strategy (**Figure 1*a***), and a different strategy is needed to achieve narrow distributions (**Figure 1*b***). Note that, without fluctuations, Equation 1 becomes $x' = x$ if $t_d = \ln 2/\lambda$, so that cell size is maintained. This example therefore

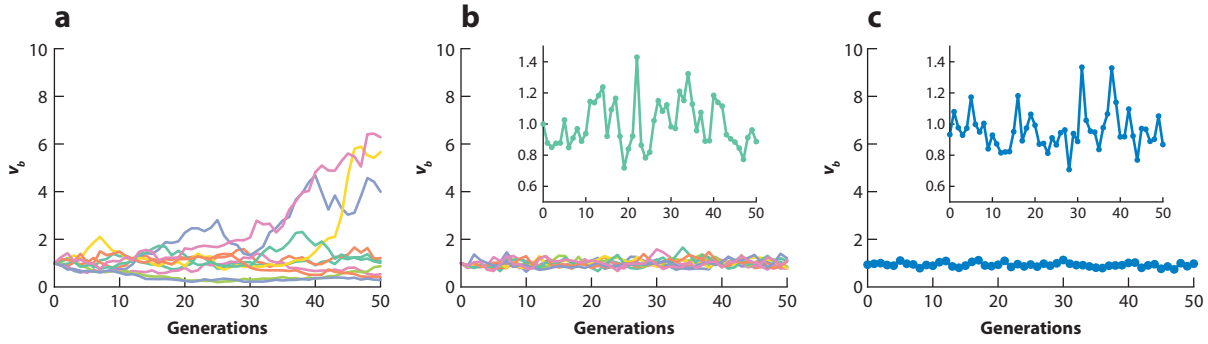

Figure 1

Cell size at birth in (*a*,*b*) simulations and (*c*) experiments as a discrete time series, with insets showing a close-up view of one particular trial of simulation (*b*) and the data (*c*). (*a*,*b*) Multiple trials (*different colors*) of numerical simulations of the discrete stochastic map in Equation 9 with the simple one-line pseudocode $v_{i+1} = (2(1-\alpha)v_i + v_0)2^{\xi_t}/2$, where v_i denotes cell size at birth in the ith generation, ξ_t is a normally distributed random variable with zero mean and variance σ_t^2, and v_0 is a constant that sets the mean cell size. In this case, we set $\sigma_t = 0.22$ and $\langle v_b \rangle = 1$. (*a*) $\alpha = 0$ leads to unconfined diffusion and a divergent distribution. (*b*) Any $0 < \alpha < 2$, in this case $\alpha = 0.5$, has the necessary feedback to achieve a stationary distribution. (*c*) v_b in this case represents cell length at birth and is normalized so that $\langle v_b \rangle = 1$. Data taken with permission from Reference 53.

shows that it is necessary to introduce stochasticity to models of cell size regulation, as the failure of a timer strategy cannot be revealed otherwise.

Third, researchers recently investigated the properties of the resulting size regulation strategy from models of molecular mechanisms that do not specify the identity of the molecular players but that nonetheless propose concrete molecular network architectures. Two models were considered, one proposing that cell division is triggered by the accumulation to a threshold number of an initiator protein (48) and another proposing that the dilution of an inhibitor triggers an event in cell cycle progression (15). It was shown theoretically that, in the context of budding yeast, *Saccharomyces cerevisiae*, both of these seemingly reasonable size regulation strategies fail to regulate cell size in the case of symmetric division (6). While there could be other explanations, this appears to be a strong constraint that may have contributed to the evolution of asymmetric division in budding yeast.

These very different examples show how phenomenological modeling, in combination with single-cell- or bulk-level experiments quantifying cell growth, can lead to biologically relevant conclusions and constrain biological mechanisms. Furthermore, this approach allows construction of a theoretical phase diagram (e.g., 34) showing not where biology lies, but where biology may exist. In this vein, we proceed with the following aphorism in mind: "All models are wrong; some models are useful" (7, p. 440].

In this review, we describe various existing phenomenological models for cell size regulation, some dating back decades and many very recent, and also present several novel results. First, we introduce discrete stochastic maps (DSMs) to model cell size regulation and the various, approximate methods of solving for the distributions and correlations they generate. We discuss the connection between the problem of cell size regulation and those of diffusion in a confining potential and autoregressive modeling in time-series analysis. We systematically show, for the first time to our knowledge, that cell size regulation in *E. coli* can be approximated well by a stochastic model where the cell size at the next generation depends only on the cell size at the present generation. Next, we review continuous rate models (CRMs) and their mapping to DSMs. We then review recent works that analyzed DSMs at higher precision. These works revealed that power-law tails may be generated by DSMs, a result not captured by the approximate methods. Finally, we review recent results beyond the phenomenological level, including molecular implementations of different strategies for cell size regulation, the problem of protein number regulation, and the effects of cell size regulation on population growth.

2. MODELS FOR CELL SIZE REGULATION

To resolve the problem of an unconfined random walk in log sizes in Equation 1, feedback must be introduced so that larger cells divide sooner than average. Some insight into the problem of cell size regulation can be gained by considering the familiar scenario of overdamped Brownian motion in a confining potential. This scenario can be described by the Langevin equation describing the dynamics of position x (16),

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{1}{\gamma}V'(x) + \sigma\xi. \qquad 2.$$

In this equation, γ is a drag coefficient that relates the force to the velocity in the overdamped limit, $V(x)$ is the confining potential, and σ is the magnitude of the fluctuations described by the stochastic variable ξ, which has correlations $\langle\xi(t')\xi(t'+t)\rangle = \delta(t)$. In this review, $\langle\cdot\rangle$ denotes the ensemble average. In the absence of a potential $V(x) = 0$, Equation 2 reduces to unconfined diffusion, the hallmark of which is the linear dependence of the mean squared displacement $\langle x^2\rangle$

on time. In a quadratic potential $V(x) = kx^2/2$, Equation 2 corresponds to diffusion confined by a linear restoring force. In this case, Equation 2 is known as an Ornstein-Uhlenbeck (OU) process and is useful in describing a plethora of physical phenomena. It can be written as

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{k}{\gamma}x + \sigma\xi, \qquad 3.$$

where k is the strength of the restoring force.

The probability density $p(x, t)$ corresponding to Equation 3 satisfies the Fokker-Planck equation that describes its temporal dynamics (16),

$$\frac{\partial p}{\partial t} = \frac{k}{\gamma}\frac{\partial}{\partial x}(xp) + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2}. \qquad 4.$$

The stationary $\partial p/\partial t = 0$ solution is a Gaussian distribution,

$$p(x) = \sqrt{\frac{k}{\pi\gamma\sigma^2}}\exp\left(-\frac{kx^2}{\gamma\sigma^2}\right). \qquad 5.$$

Indeed, Equation 5 is equal to the Boltzmann distribution $p(x) \propto \exp[-V(x)/k_\mathrm{B}T]$, where k_B is the Boltzmann constant and T is the temperature, since $\sigma^2 = 2D = 2k_\mathrm{B}T/\gamma$ by the Einstein relation for the diffusion coefficient D. As *karro w*0, the strength of the confining potential weakens. At $k = 0$, the variance of x diverges. However, for any $k > 0$, the variance of x will be finite. The autocovariance $\langle x(t')x(t'+t)\rangle$ can be obtained via integration of Equation 3. At stationarity, the autocovariance is exponentially decaying (16):

$$\langle x(t')x(t'+t)\rangle = \frac{\gamma\sigma^2}{2k}\exp\left(-\frac{k}{\gamma}|t|\right). \qquad 6.$$

The familiar example of an OU process turns out to be similar to the problem of cell size regulation, but with the variable x now representing cell size. In the following sections, we review the formulation of the problem of cell size regulation as a discrete analog of an OU process.

2.1. Discrete Stochastic Maps

Figure 1*c* shows single-cell data obtained via microfluidic devices that trap single cells in microchannels to allow measurements of physiological properties such as cell size for many generations (52, 53, 59). The problem of cell size regulation may be investigated initially by consideration of only division events. The data in this case consist of cell size at birth, division, and interdivision times over many generations. What are the distributions and correlations of cell size at birth and division, and what size regulation strategies lead to such statistics?

At a phenomenological, coarse-grained level, a size regulation strategy can be specified as a map that takes cell size at birth v_b to a targeted cell size at division v_a with a deterministic strategy $f(v_b)$ (3),

$$v_a = f(v_b). \qquad 7.$$

In the face of biological stochasticity, the actual cell size at division v_d is v_a, subject to some coarse-grained noise term. For example, the noise term can be size additive, so that $v_d = v_a + \xi_v$, where ξ_v is uncorrelated between generations. The noise term can also be time additive. In this case, the stochastic interdivision time t_d can be written as $t_d = t_a + \xi_t$, where ξ_t is the noise term. The deterministic component t_a can be determined by assuming a constant exponential growth rate λ. The deterministic size regulation strategy in Equation 7 then leads to

$$t_a = \ln(f(v_b)/v_b)/\lambda. \qquad 8.$$

In the case of time-additive noise, if division is perfectly symmetric, so that $v_b' = v_d/2$, then the cell size at birth at the next generation is

$$v_b' = f(v_b)e^{\lambda\xi_t}/2. \quad 9.$$

The two forms of noise lead to distributions of different shapes. Experiments have shown that distributions of cell sizes at birth are skewed and can be approximated as a log-normal, but that interdivision time distributions can be approximated as normal (11, 58). These results are consistent with a normally distributed time-additive noise, which we use below.

2.2. Approximate Solution via First-Order Expansion

The DSM described in Section 2.1 is, in general, difficult to solve for an arbitrary size regulation strategy $f(v_b)$. One solution method makes the approximation to focus on the behavior of f near the mean size $\langle v_b\rangle$, since the size distribution has a small CV. A size regulation strategy can be linearized by expanding about $\langle v_b\rangle$, $f(v_b) \approx f(\langle v_b\rangle) + f'(\langle v_b\rangle)(v_b - \langle v_b\rangle)$. In this approximation, all regulation strategies that agree to first order will lead to similar distributions near $\langle v_b\rangle$. The following is a convenient choice (3):

$$f(v_b) = 2v_b^{1-\alpha}v_0^{\alpha}, \quad 10.$$

where v_0 is an arbitrary constant. As shown below, we find that $\langle v_b\rangle \approx v_0$ and thus that the slope has value $f'(\langle v_b\rangle) = 2(1-\alpha)$. The value of α therefore determines the strength of regulation. $\alpha = 1$ corresponds to the strongest regulation, a sizer strategy where cells attempt to divide upon reaching $f(v_b) = 2v_0$. $\alpha = 0$ represents no regulation and corresponds to the timer strategy, where cells attempt to divide upon reaching $f(v_b) = 2v_b$. Recent works have shown that the statistics of cell size can be generated by a regulation strength $\alpha = 1/2$ that is between the two extremes (11, 44, 51, 58). In this case, the slope has value $f'(\langle v_b\rangle) = 1$ and thus is an approximation to the adder strategy [also known as the incremental model (56, 57)], where cells attempt to divide upon reaching $f(v_b) = v_b + v_0$. Several microorganisms in all three domains of life have been shown to approximately follow an adder strategy or, less prescriptively, to exhibit adder correlations. We discuss the prevalence of adder correlations below.

Let $x = \ln(v_b/v_0)$ be the log size and x' denote x at the next generation. Equations 9 and 10 then lead to the simple stochastic equation

$$x' = (1-\alpha)x + \lambda\xi_t. \quad 11.$$

The log size at the nth generation can be written as

$$x_n = (1-\alpha)^n x_0 + \sum_{i=0}^{n-1}(1-\alpha)^{n-1-i}\lambda\xi_t^{(i)}, \quad 12.$$

where x_n and $\xi_t^{(i)}$, respectively, denote the value of x and ξ_t at the nth generation. The first term approaches zero as $narrow\infty$ if $0 < \alpha < 2$. If ξ_t is normally distributed with variance σ_t^2, then the variance σ_x^2 of x will be the sum of the variances in the series in the second term. The geometric series converges for $0 < \alpha < 2$ and can readily be evaluated to give the variance σ_x^2 as

$$\sigma_x^2 = \frac{\lambda^2\sigma_t^2}{\alpha(2-\alpha)}. \quad 13.$$

Furthermore, since x_n is a sum of normal variables, it will also be normally distributed. If $\alpha \leq 0$ or $\alpha \geq 2$ hold true, then the sum of the series diverges and, thus, there is no stationary distribution. The case $\alpha = 0$ produces unconfined diffusion and is analogous to the case where the strength

of the restoring force is zero ($k = 0$) in an OU process, as seen in Equation 5. **Figure 1*a*,*b*** demonstrates the difference between time series generated by $\alpha = 0$ and by $0 < \alpha < 2$. The variance of t_d can be obtained similarly.

It is not obvious a priori whether the widths of the distributions of interdivision time and cell size are related. It turns out that the two CVs [denoted by CV(·)] are related by a dimensionless quantity (3). The log size is related to the actual size by $x = \ln(v_b/v_0) \equiv \ln(1 + \delta v_b)$. Since $\delta v_b = v_b/v_0 - 1$ is small, the log-size can be approximated as $x \approx \delta v_b = v_b/v_0 - 1$. Therefore, we find that $\mathrm{CV}(v_b) \approx \sigma_x$. Calculating $\mathrm{CV}(t_d)$ in a similar manner leads to

$$\frac{\mathrm{CV}(v_b)}{\mathrm{CV}(t_d)} \approx \frac{\ln 2}{\sqrt{2\alpha}}. \quad 14.$$

Equation 14 allows extraction of the parameter α from CVs that can be accurately measured. Since both x and t_d are distributed normally, the model predicts that these distributions can be collapsed after normalizing by the mean and scaling according to Equation 14, as has been demonstrated in experiments (47, 51).

One can also obtain the Pearson correlation coefficients (CCs) between two variables [denoted by $C(\cdot, \cdot)$]. Since CCs are not affected by addition or multiplication by a constant, v_b can be replaced by x in the following calculations. The CC between the cell size at birth of a mother cell and that of the daughter cell is therefore (3)

$$C(v_b, v_b') = C(x, x') = \frac{\langle xx' \rangle - \langle x \rangle^2}{\sigma_x^2}. \quad 15.$$

Substituting in Equation 11, we find that

$$C(v_b, v_b') = 1 - \alpha. \quad 16.$$

Importantly, the value of $\alpha \approx 1/2$ extracted via the ratio of CVs in Equation 14 also predicts the CC between size at birth and at division, in agreement with experimental results (11, 30, 51).

Similarly, using Equations 8 and 10, the interdivision time can be written as

$$t_d = \frac{\ln 2 - \alpha x}{\lambda} + \xi_t. \quad 17.$$

The CC between the interdivision times of a mother–daughter pair can then be shown to be (32, 51)

$$C(t_d, t_d') = -\alpha/2. \quad 18.$$

The fact that this CC is nonzero has implications for the population growth rate, which we review below.

2.3. Autoregressive Models and Extensions to Incorporate Biological Details

Equation 11, obtained after linearization of the generically nonlinear DSM in Equation 9, is mathematically known as an autoregressive (AR) model, often used in time-series analysis and economics forecasting (42). An AR model of order m [denoted by AR(m)] takes the form $x_i = b + \sum_{j=1}^{m} c_j x_{i-j} + \xi_i$, where b and c_j are constants and ξ_i is a noise term uncorrelated for different i. The model describes how previous values of the stochastic variable x linearly influence the next value. Equation 11 is an AR(1) model, which is also a discrete analog of an OU process.

In the problem of cell size regulation, it is not obvious a priori if an AR(1) model is sufficient to describe data. One method to determine the appropriate order of an AR model is to investigate the partial CCs. The partial CC between x_i and x_{i+2} given an intermediate variable x_{i+1} is defined

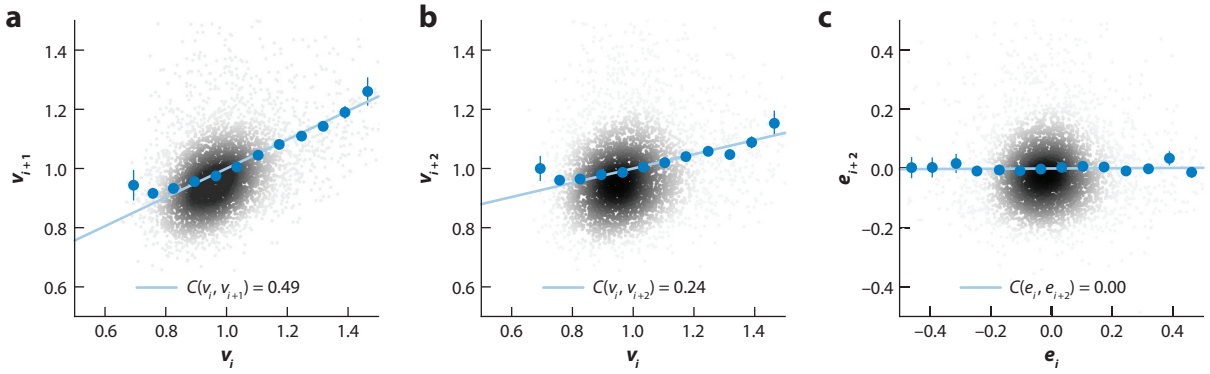

Figure 2

Cell size at birth in *Escherichia coli* can be described by an autoregressive model of order 1. (*a*,*b*) Single-cell data of cell size at birth v_i at the ith generation and the resulting correlation coefficients (CCs) between (*a*) parent and children and (*b*) parent and grandchildren. (*c*) The residuals e_i after linear regression with v_{i+1} and the resulting CC, $C(e_i, e_{i+2})$. Gray dots show data, with shading representing the density of points. Blue circles show average values binned according to values on the x axis. Error bars show one standard error of the mean. Blue lines show the best linear regression of raw data. Data are taken from Reference 53.

as the CC between the residuals $e_i = x_i - \hat{x}_i$ and $e_{i+2} = x_{i+2} - \hat{x}_{i+2}$. $\hat{x}_i = p_1 x_{i+1} + p_0$ and $\hat{x}_{i+2} = q_1 x_{i+1} + q_0$ denote the predicted value after linear regression, with x_{i+1} determining the coefficients. The resulting partial CC between x_i and x_{i+2}, with the intermediate variable x_{i+1}, is (42)

$$C_{x_{i+1}}(x_i, x_{i+2}) = C(e_i, e_{i+2}) = \frac{C(x_i, x_{i+2}) - C(x_i, x_{i+1})C(x_{i+1}, x_{i+2})}{\sqrt{\left[1 - C^2(x_i, x_{i+1})\right]\left[1 - C^2(x_{i+1}, x_{i+2})\right]}}. \quad 19.$$

In an AR(1) model, the CC $C(x_i, x_{i+2})$ is nonzero because x_i and x_{i+2} are related via the intermediate variable x_{i+1}. However, the partial CC $C_{x_{i+1}}(x_i, x_{i+2})$ removes the effects of the intermediate variable and is zero. Experimentally determined values of $C(v_i, v_{i+1})$ are as predicted by Equation 16 (**Figure 2*a***) (53). In the same data set, $C(v_i, v_{i+2})$ is nonzero, but $C_{v_{i+1}}(v_i, v_{i+2})$ is zero (**Figure 2*b*,*c***). Indeed, a vanishing $C_{v_{i+1}}(v_i, v_{i+2})$ implies that $C(v_i, v_{i+2}) = C^2(v_i, v_{i+1})$ holds true, which is the case. This novel check systematically shows that cell size at birth in *E. coli* can be described by an AR(1) model. This result is a fortunate simplification since, for example, in certain mammalian cells, the CCs in the interdivision times between cousin cells (C_{cc}) cannot be determined from those between sister cells (C_{ss}) and between cells in mother–daughter pairs (C_{md}). Instead, experiments observe that $C_{cc} > C_{md}$, contrary to the expected relationship $C_{cc} = C_{md}^2 C_{ss}$ in an AR(1) model (37, 43).

The parameters of an AR model can be estimated by relating their values to the theoretical values of the autocorrelation function (ACF) (42), as is done in Equation 16 to extract the regulation strength α. The ACF $\rho(t)$ is the CC between variables separated by t time points,

$$\rho(t) = C(x_i, x_{i+t}). \quad 20.$$

As can be seen in Equation 12, the ACF for the AR(1) model of Equation 11 is simply

$$\rho(t) = (1 - \alpha)^{|t|}. \quad 21.$$

In this case, the ACF decays exponentially, as in an OU process, as seen in Equation 6. The ACF of cell size at birth indeed decays exponentially (**Figure 3*a***) (54). Importantly, the estimated ACF is only meaningful after sufficient averaging to eliminate spurious fluctuations. This can be done

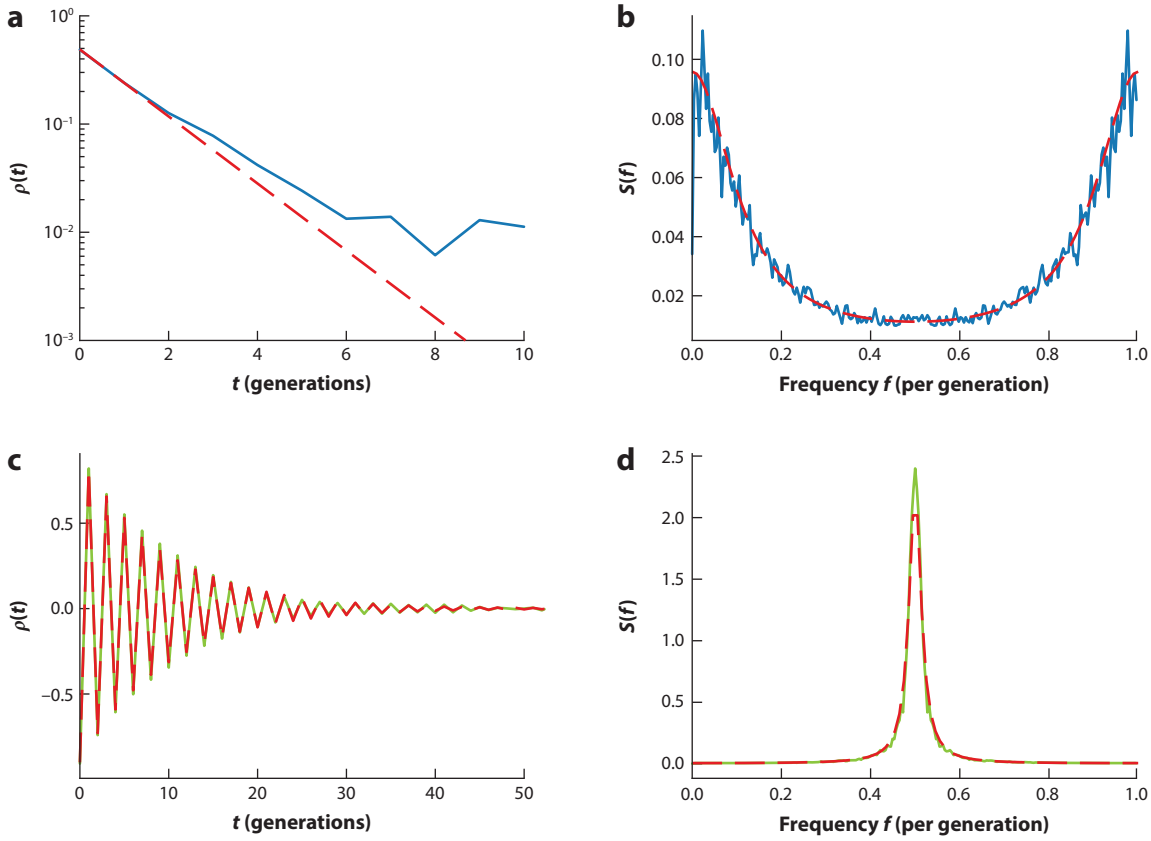

Figure 3

Fluctuations versus oscillations. (*a*,*c*) The autocorrelation function (ACF) $\rho(t)$ and (*b*,*d*) the power spectral density (PSD) $S(f)$ (estimated via the Welch method) of cell size at birth in (*a*,*b*) *Escherichia coli* and (*c*,*d*) a simulated autoregressive model of order 1, described by Equation 11, with $\alpha = 1.9$. Blue lines show experimentally determined ACF and PSD from the data set in Reference 53. Green lines show simulation results. Dashed red lines show (*a*,*c*) Equation 21 and (*b*,*d*) Equation 23 for (*a*,*b*) $\alpha = 0.49$ and (*c*,*d*) $\alpha = 1.9$.

most effectively by computing the power spectral density (PSD),

$$S(f) = \lim_{T \to \infty} \frac{1}{T} \left| \sum_{j=1}^{T} x_j e^{-i2\pi fj} \right|^2, \qquad 22.$$

where T is the total number of observations in the time series with data points x_j. The PSD can also be calculated as the Fourier transform of the ACF according to the Wiener-Khinchin theorem. For the AR(1) model of Equation 11, the PSD turns out to be (42)

$$S(f) = \frac{\alpha(2-\alpha)}{1 - 2(1-\alpha)\cos(2\pi f) + (1-\alpha)^2} \sigma_x^2. \qquad 23.$$

This is again analogous to an OU process, since the Fourier transform of an exponential function is a Lorentzian function. There are significant oscillations only in the case $\alpha \lesssim 2$, for which the PSD peaks at high frequencies (**Figure 3*c*,*d***). The case of *E. coli*, where $\alpha \approx 1/2$, is far from this regime (**Figure 3*a*,*b***).

The AR(1) model in Equation 11 can be extended to incorporate details that are relevant to a variety of microorganisms. These include asymmetric and noisy divisions [e.g., in mycobacteria (2, 33)], noisy growth rates [e.g., in slow-growing *E. coli* (58) and in the archaeon *Halobacterium salinarum* (14)], and diverse growth morphologies [e.g., the budding mode of growth of *S. cerevisiae* (47)]. Noisy divisions and noisy growth rates can be incorporated by modeling the division ratio (daughter cell size at birth divided by mother cell size at division) and growth rate as $1/2 + \xi_r$ and $\lambda + \xi_\lambda$, respectively, at each generation. If the fluctuations ξ_r and ξ_λ are small and uncorrelated, Equation 11 becomes, to first order in small variables,

$$x' \approx (1-\alpha)x + \lambda\xi_t + 2\xi_r. \quad 24.$$

The fluctuation ξ_λ enters as a first-order correction to interdivision time t_d. The CVs and CCs can be calculated as before in Equation 24 to show that the different fluctuations typically affect the CVs and CCs in different ways. For example, the CC between cell size at birth and at division, $C(v_b, v_d)$, is sensitive to fluctuations in division ratios and is increased by large fluctuations in division ratios. However, the CC in cell size at birth between cells in mother–daughter pairs, $C(v_b, v_b')$, remains the same as in Equation 16 and is independent of all noise terms. It is thus a robust detector of the underlying regulation strategy even in the face of multiple sources of complicating stochasticity (14). We discuss below several models that incorporate other biological details and move beyond AR models.

2.4. Continuous Rate Models and Higher-Order Effects

Cell size regulation can also be modeled using CRMs (20, 27, 38, 51). In contrast to DSMs, CRMs consider not just discrete division events, but also the continuous cell cycle. They specify the instantaneous division rate h, or the probability to divide per unit size increment, as a function of physiological parameters such as the current size v, size at birth v_b, growth rate λ, or the time t since division. A simple choice of parameterization is the sloppy sizer model, $h = h(v)$. In this case, the probability for a cell of size v to divide between the size intervals v and $v + \mathrm{d}v$ is $h\,(v)\,\mathrm{d}v$. Thus, if $F(v_d|v_b)$ is the cumulative probability that a cell has not divided at size v_d given v_b, then $F(v_d|v_b)$ satisfies $F(v_d + \mathrm{d}v|v_b) = F(v_d|v_b)\,[1 - h(v_d)\mathrm{d}v]$. In the continuum limit, the above expression becomes $\mathrm{d}F(v_d|v_b)/\mathrm{d}v = -h(v_d)\,F(v_d|v_b)$. Therefore, $F(v_d|v_b)$ satisfies

$$F(v_d|v_b) = \exp\left[-\int_{v_b}^{v_d} h(v)\mathrm{d}v\right]. \quad 25.$$

Equation 25 can be written as

$$h(v) = -\frac{\mathrm{d}}{\mathrm{d}v}\ln F(v|v_b), \quad 26.$$

allowing one to extract $h(v)$ via single-cell experiments that measure $F(v_d|v_b)$. The division rate can be formulated as a probability that the cell divides per unit time increment, as well, using the change of variables between size and time given by exponential growth. Analyses using CRMs have demonstrated that the current size is not the only determinant of the division rate because the sloppy sizer model fails to capture measured distributions of interdivision time and size increment from birth to division (38). This implies that there exists a feedback on the time since birth, or equivalently, on the size at birth (38). Specifically, a division rate in the form $h = h(v - v_b)$ can simultaneously describe measured distributions of size at birth, interdivision time, and size increment from birth to division (51).

A CRM can be approximately reduced to a DSM with the target size at division equal to the expectation value of the size at division given the size at birth (20):

$$f(v_b) = \int_0^\infty p(v|v_b) v \mathrm{d}v, \qquad 27.$$

where $p(v|v_b) = -\mathrm{d}F(v|v_b)/\mathrm{d}v$ is the probability density for a cell born at size v_b to divide at size v. The nature and magnitude of the noise term can be determined by inverting the steps described below to map a DSM to a corresponding CRM. To do so, $p(v|v_b)$ can be calculated from $f(v_b)$ and a specified coarse-grained noise, and the division rate can then be obtained using Equation 26. For example, for a time-additive, normally distributed noise with variance σ_t^2, the division probability density for log size $x = \ln(v/v_0)$ is $p(x|x_b) \propto \exp\{-[x - g(x_b)]^2/[2\lambda^2\sigma_t^2]\}$, where $g(x_b) = \ln[f(v_0 e^{x_b})/v_0]$. Since the typical x_b is much smaller than $g(x_b)$, integration leads to the division rate (20)

$$h(v, v_b) \approx \frac{\sqrt{2}}{v\sqrt{\pi}\lambda\sigma_t} H\left[\frac{1}{\sqrt{2}\lambda\sigma_t} \ln\left(\frac{v}{f(v_b)}\right)\right], \qquad 28.$$

where $H(z) = \exp(-z^2)/[1 - \mathrm{Erf}(z)]$ and $\mathrm{Erf}(\cdot)$ is the error function. For the regulatory function in Equation 10, the division rate in Equation 28 becomes a function of only the instantaneous size when $\alpha = 1$, corresponding to a sizer strategy.

Although the CRM is generic and may capture complex behavior such as filamentation (38), it is not obvious a priori how to parameterize the division rate. In contrast, the DSM has only a few parameters, is amenable to analytical treatment in several cases, and describes existing measurements well. The complexity sacrificed by DSMs and their first-order approximate solutions may become important, for instance, when second- and higher-order terms become significant. However, higher-order effects are difficult to detect unless the number of cells measured is large enough to suppress the confounding effects of fluctuations in the cell cycle. No existing experiments have achieved this, perhaps justifying the success of DSMs as models of cell size regulation (20).

2.5. More Precise Analyses of Discrete Stochastic Maps

The approximate first-order solution in Section 2.2 predicts a log-normal size distribution for a regulatory function $f(v_b) = 2(1 - \alpha)v_b + v_0$ and a time-additive, normally distributed noise. However, closer inspection reveals that the size distribution has a power-law tail instead. This can be seen by analyzing which moments exist for a given regulation strength α. Calculations similar to those in Section 2.2 show that if the jth moment exists, so too do all the lower moments, but that for any $\alpha > 0$, there always exists an integer j^* past which all moments cease to exist. This suggests that the size distribution has a power-law tail $p(v_b) \sim 1/v_b^{1+\beta}$ with $j^* < \beta \leq j^* + 1$, as confirmed by numerical simulations (34).

The value of β can be obtained precisely. The evolution of size distributions from one generation to the next can be written as an integral equation $p(v_b') = \int_0^\infty K(v_b', v_b) p(v_b) \mathrm{d}v_b$, where the kernel $K(v_b', v_b)$ can be derived from the regulatory strategy $f(v_b)$. For a regulatory function in the form $f(v_b) = 2(1 - \alpha)v_0(v_b/v_0)^\eta + v_0$, it can be shown via an asymptotic analysis of the integral equation that a distribution with a power-law tail $1/v_b^{1+\beta}$ evolves to one with a power-law tail $1/v_b^{1+\beta/\eta^2}$ (34). This implies that, for $\eta = 1$, the stable size distribution indeed has a power-law tail, with

$$\beta = \frac{-2\ln(1 - \alpha)}{\lambda^2\sigma_t^2}, \qquad 29.$$

where σ_t^2 is the variance of the time-additive noise.

An alternative approach also led to the same power-law tail (28). In this approach, a DSM is approximated as a Langevin equation continuous in generations. Let $x = \ln(v_b/v_0)$ be the log size at birth, and let n denote the generation number. In this case, Equation 9 can be written

$$x_{n+1} = x_n + \tilde{g}(x_n) + \lambda\xi_t, \qquad 30.$$

where $\tilde{g}(x_n) = \ln\{f[\exp(x_n)v_0]/v_0\} - x_n$. To the lowest order, Equation 30 can be approximated by a Langevin equation continuous in n as

$$\frac{\mathrm{d}x}{\mathrm{d}n} = \tilde{g}(x) + \lambda\xi_t. \qquad 31.$$

As seen above, an OU process, described by Equations 3 and 31, leads to an equilibrium distribution of log sizes $p(x) \propto \exp[-2V(x)/(\lambda^2\sigma_t^2)]$, where $V(x) = \int \tilde{g}(x')\mathrm{d}x'$ is the effective potential. For the same regulatory function as above, the effective potential diverges linearly as $V(x) \sim -2x\ln(1-\alpha)$. The equilibrium distribution therefore has a power-law tail $1/v_b^{1+\beta}$ with the same β as in Equation 29 (28). Even further precision can be obtained via a second-order approximation that modifies the effective potential but leaves the behavior of the power-law tail unchanged (28).

3. BEYOND PHENOMENOLOGICAL MODELS OF CELL SIZE REGULATION

As we alluded to above, the formalism of DSMs developed for the problem of cell size regulation can lead to insights on related problems at the molecular, single-cell, and population levels. In the sections below, we discuss each of these levels in turn.

3.1. Molecular Mechanisms to Implement Cell Size Regulation

How does a bacterial cell molecularly implement a size regulation strategy? The initiator accumulation model is a network architecture proposing that an initiator protein accumulates during cell growth to trigger cell division upon reaching a threshold copy number θ (48). While experiments have suggested that the upstream control occurs over initiation of DNA replication rather than cell division in various microorganisms (4, 13, 47, 62), we first review a simpler model where the accumulation of initiators triggers cell division. This model leads to the adder correlations observed in several species of bacteria and other microorganisms (3, 6, 24).

One possible molecular implementation of the initiator accumulation model is as follows (17): If the transcription rate of the initiator is assumed to be proportional to the cell volume, which grows exponentially in time, and if each transcript leads to a burst of protein production with mean burst size b, then the distribution of added cell size $\Delta v = v_d - v_b$ from birth to division has width (17)

$$\mathrm{CV}^2(\Delta v) = \frac{b^2 + 2b\theta + \theta}{(b+\theta)^2}. \qquad 32.$$

Furthermore, the resulting distribution has only one characteristic size, the mean added cell size $\langle\Delta v\rangle$, and therefore can be written as

$$p(\Delta v) = \frac{1}{\langle\Delta v\rangle}\tilde{p}\left(\frac{\Delta v}{\langle\Delta v\rangle}\right). \qquad 33.$$

Indeed, experiments showed that distributions of cell sizes with different means collapse after normalizing by the mean (18, 27, 51, 55). The collapse suggests that b and θ are constant within the implementation in this case. The same experiments also showed that the distributions of

interdivision times collapse after normalizing by the mean doubling time, which is again captured by this model (17). There are also additional models that show such scaling collapse, such as an autocatalytic network subject to a threshold criterion for division (25) and the coarse-grained adder-per-origin model described below.

As discussed in Section 1, control at other cell cycle events may lie upstream of cell division in various microorganisms. DSMs similar to those reviewed above can be extended to describe cell cycle regulation. These models can not only produce emergent strategies of cell size regulation identical to those described by the division-centric models reviewed above, but can also describe additional statistics such as the correlations between cell size and various cell cycle timings (1, 39).

As an example, we review a model of cell cycle regulation in *E. coli*, whose cell divisions appear to follow a constant time T after the initiation of DNA replication for a broad range of mean growth rates (12, 60). The time T can be larger than the mean doubling time τ, in which case the cells maintain multiple ongoing rounds of DNA replication. The tight coupling between initiation and division implies that the cell size at birth v_d is

$$v_d = v_i e^{\lambda(T+\xi_T)}, \qquad 34.$$

where v_i is the cell size at initiation, $\lambda = \ln 2/\tau$ is the growth rate, and ξ_T describes fluctuations with magnitude σ_T in the time between initiation and division. At a coarse-grained level, the initiator accumulation model can be described as (3, 24, 48)

$$\tilde{v}_i' = (v_i + Ov_0)e^{\lambda\xi_t}, \qquad 35.$$

where $\tilde{v}_i'$ is the total cell size of the daughter cells (typically two) at the next initiation, O is the number of origins of replication (i.e., the site along the chromosome at which DNA replication initiates), and v_0 is a constant. As in the division-centric model, regulation is subject to a time-additive noise ξ_t with magnitude σ_t.

Analysis and simulations of the initiation-centric model of Equations 34 and 35 show that it produces emergent adder correlations at division (6, 24) as long as the magnitude of the fluctuations in the coordination between initiation and division is much less than that in the control of initiation ($\sigma_t \gg \sigma_T$). This is indeed the case in experiments for fast-growing bacteria, although the picture appears different for slow-growing bacteria (58), which we discuss below. The model also generates cell size and interdivision time distributions, the CVs of which depend only on the magnitudes $\lambda\sigma_t$ and $\lambda\sigma_T$ of the fluctuations and the regulation strength α. The distributions therefore collapse after scaling by the mean if these parameters are constant across growth conditions. At the bulk level, the initiation-centric model produces the observed exponential scaling of mean cell size with mean growth rate, as discussed in Section 1, without requiring parameters to depend on mean growth rate (24). These results, together with previous results regarding the universality of cell size distributions, suggest that the initiator accumulation model may be a robust molecular mechanism that produces adder correlations, and that models of cell cycle regulation can continue to shed light on the underlying biology.

3.2. Regulation of Protein Numbers

Recent works have begun investigating the statistics of the copy numbers of proteins at the single-cell level in the same spirit as the problem of cell size regulation (9, 10, 50, 54). In fact, for a constitutively expressed protein, the distributions of protein numbers at birth can be described by a DSM (10). Analysis analogous to that in **Figure 2**, but applied to the copy number of a constitutively expressed protein in *E. coli* in the same data set (53), reveals that the partial CC is also zero in this case. However, it is unclear how protein number and cell size are simultaneously regulated.

One way to investigate this question is via a multidimensional, or vector, AR model. An AR(1) vector model in M dimensions can be written as

$$\mathbf{x}' = A\mathbf{x} + \mathbf{b} + \boldsymbol{\xi}, \quad 36.$$

where $\mathbf{x}$ is a vector of the abundances at birth of the M cellular components, which can include cell size, and $\mathbf{x}'$ is the vector at the next generation. A is an $M \times M$ matrix representing the regulatory interactions between components, $\mathbf{b}$ is a vector representing the basal synthesis level between generations, and $\boldsymbol{\xi}$ is a vector of noise terms that are uncorrelated between generations but may be cross-correlated at the same generation.

For the one-dimensional case, the condition for stationarity is that $2 > \alpha > 0$ so that the variance of x in Equation 13 is finite. This condition is equivalent to that in which the zero of $1 - (1 - \alpha)z$ lies outside the unit circle. In the multidimensional case, the condition is, similarly, that all of the zeros of $\det(I - Az)$ lie outside the unit circle (42). Given a stable AR(1) vector model, the multidimensional analog of Equation 16 can be used to estimate the regulatory matrix A from measurements (8). For the data set discussed above (53), this method results in

$$A = \begin{pmatrix} 0.50 \pm 0.02 & -0.02 \pm 0.01 \\ -0.16 \pm 0.02 & 0.60 \pm 0.02 \end{pmatrix},$$

where the first and second components are, respectively, cell size and protein number of the constitutively expressed protein at birth (both normalized by their means), and plus–minus shows the standard error in the estimate. This novel result suggests that the copy number of this constitutively expressed protein does not affect cell size regulation, while cell size does affect the regulation of this protein number. It is unknown whether this result holds for all constitutively expressed proteins and how this result will change for proteins that are not constitutively expressed.

To better understand cross-correlations between cell size and protein numbers from a mechanistic perspective, recent works have investigated a dynamical model in the form $\mathrm{d}\mathbf{x}/\mathrm{d}t = A\mathbf{x}$, where $\mathbf{x}$ is now the abundances of the cellular components during the cell cycle, and A now describes the regulatory interactions in time (50). This model leads to the components growing as a sum of exponentials that can be approximated as a single exponential function during one generation, in agreement with experimentally observed exponential growth (9). Describing the statistics generated by dynamical models and relating a dynamical model to a DSM and vice versa remain important open research areas.

3.3. Effects of Cell Size Regulation on Population Growth Rate

At the single-cell level, genetically identical cells in the same clonal populations may have different interdivision times and growth rates. How does such variability at the single-cell level affect population growth? Models often assume that the interdivision times t_d are uncorrelated between generations and independent of other variables (22, 26, 41). In this case, a simple relation connects the asymptotic population growth rate $\Lambda = (\mathrm{d}N/\mathrm{d}t)/N$, where N is the number of cells in the population, to the interdivision time distribution $p(t_d)$:

$$2\int_0^\infty p(t_d)\exp(-\Lambda t_d) = 1. \quad 37.$$

Importantly, given a fixed mean interdivision time, a larger variability in t_d increases Λ. However, cell size regulation leads to negative correlations in t_d between generations, as seen in Equation 18. In this case, recent results showed that, in an asynchronous, exponentially growing population—in which each cell is subject to variability in its single-cell growth rate, as well as to time-additive and

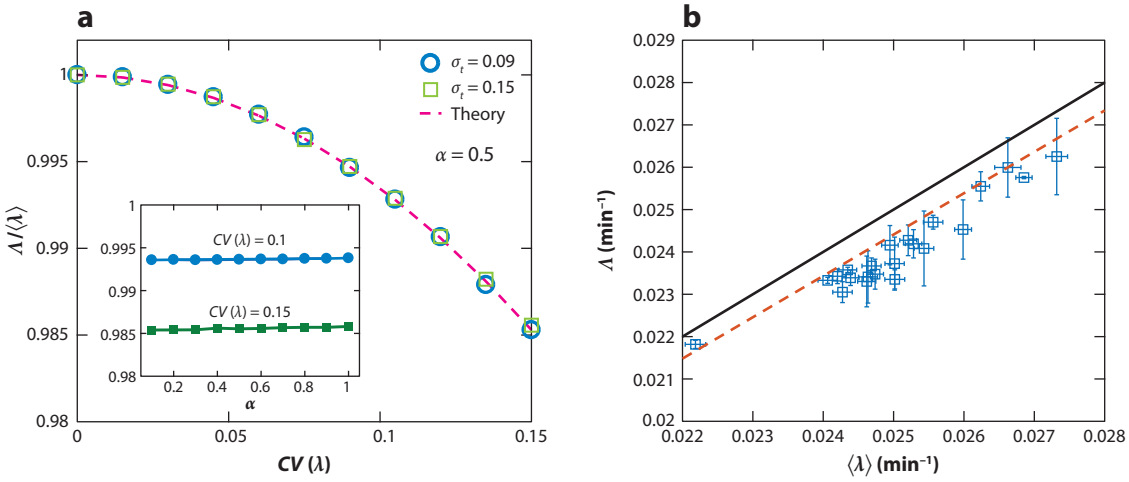

Figure 4

Cell size regulation, as long as it exists, does not affect the population growth rate. (*a*) Population growth rates obtained from simulations (*symbols*) of an exponentially growing population subject to variability in single-cell growth rates are consistent with Equation 38 (*dashed line*). The inset shows that population growth rates do not vary with the regulation strength as long as $1 > \alpha > 0$ holds true. (*b*) Variability in single-cell growth rates decreases the population growth rate. Blue squares show data from Reference 49 (details of the error bars can be found in Reference 32). The red dashed line shows Equation 38 for CV(λ), measured by experiments. The black solid line shows $\Lambda = \langle\lambda\rangle$ as a guide. Figure adapted with permission from Reference 32.

size-additive noise in its cell size regulation by the regulatory function $f(v_b) = 2(1-\alpha)v_b + 2\alpha v_0$—the population growth rate is dependent only on the distribution of single-cell growth rates. In the limit of small correlations in growth rates between generations, variability in single-cell growth rates does not increase, but rather decreases the population growth rate (32),

$$\Lambda/\langle\lambda\rangle = 1 - \left(1 - \frac{\ln 2}{2}\right) \mathrm{CV}^2(\lambda). \qquad 38.$$

Equation 38 predicts that a population can enhance its population growth rate by suppressing the variability in single-cell growth rates given a fixed mean, which is consistent with a CV of single-cell growth rates that is smaller than that of interdivision times, as observed in experiments (**Figure 4*a*,*b***) (51, 58). Equation 38 holds for any size regulation strategy $1 > \alpha > 0$, implying that cell size regulation, as long as it exists (in particular, $\alpha \neq 0$ leads instead to Equation 37), does not affect the population growth rate within the models studied in this review (**Figure 4*a***).

4. DISCUSSION

In this review, we summarize the mathematical formulations of the problem of cell size regulation, with a focus on coarse-grained, discrete models. The various models are summarized in **Table 1**. As an example, we show that a first-order AR model can describe the statistics of cell size in *E. coli*. We discuss how detailed analyses of such models led to several biologically relevant insights at the molecular, single-cell, and population levels. The same approach may shed light on several outstanding questions.

First, the prevalence of adder correlations in all three domains of life [e.g., the prokaryote *E. coli* (11, 51, 58), the eukaryote *S. cerevisiae* (47), and the archaeon *H. salinarum* (14)] suggests that an adder strategy may be simpler to implement or may be evolutionarily advantageous compared

Table 1 A summary of various models of cell size regulation reviewed here

Level	Model	Section(s)	Description	Parameters	Testable predictions	Biological insights[a]	Related works[a]
Single-cell	Discrete stochastic maps (DSMs)	2.1–2.3, 2.5	Discrete analog of random walk in confining potential	Strength of regulation	Coefficient of variations (CVs) Correlation coefficients (CCs) Autocorrelation function (ACF)	Both the cell size and the interdivision time distributions are controlled by a single mechanism	3, 54
	Continuous rate models (CRMs)	2.4	Describes cell division via a continuous function	Instantaneous division rate function	CVs, CCs, ACF	May capture complex behavior such as filamentation	38, 51
	Multidimensional DSMs	3.2	Dimensions may correspond to cell size or protein numbers	Matrix describing the regulatory interactions between the components	Cross-correlations between cell size and protein numbers	Cell size regulation may affect protein number regulation	9, 50
Molecular	Initiator accumulation	3.1	Molecular implementation of a cell size regulation strategy	Number of initiators required	CVs, CCs, ACF	Initiator accumulations robustly implements an adder size regulation strategy	24, 48
Population	Population with cell size regulation	3.3	Exponentially growing population with cell size regulation given by a DSM	Distribution of single-cell growth rates	Population growth rate	Cell size regulation does not affect population growth rate within existing models	32, 41

[a]The columns "Biological insights" and "Related works" point out only a few of the many examples in the literature.

to other size regulation strategies. An explanation of the prevalence of adder correlations remains missing, however, since cell size regulation was found not to affect the population growth rate within the class of phenomenological growth models reviewed above (32).

In contrast, the mean single-cell growth rate affects cell size regulation, since *E. coli* in slow growth conditions no longer exhibits adder correlations (58). This observation may be explained by the introduction of stochasticity in single-cell growth rates into the initiation-centric model discussed in Section 3.1, which can lead to a size regulation strategy that varies with mean growth rates (58). Indeed, cell size regulation may potentially be an emergent property of cell cycle regulation (4). This view is further supported by several models that describe cell division as a downstream effect of another cell cycle event [e.g., initiation of DNA replication in *Mycobacterium smegmatis* (33), the onset of budding in *S. cerevisiae* (6, 47), and septum constriction in *Caulobacter crescentus* (5)], and nonetheless reproduce the observed statistics at divisions.

Models of cell size regulation may also incorporate diverse growth morphologies. For example, in *S. cerevisiae*, division asymmetry depends on the duration of the budded phase, during which all cell growth occurs for the budded daughter cell (47). In *M. smegmatis*, cell growth occurs at the two poles of the cell: The old pole grows faster than the new pole, and on average, the daughter that inherits the old pole is larger (2). In both cases, the subpopulations formed by the larger and smaller daughter cells exhibit different emergent size regulation strategies (33, 47). Models incorporating these details move beyond AR models but remain straightforward

to simulate numerically, allowing the statistics they generate to be compared to experiments to distinguish between competing models.

At the molecular level, the behavior of molecular network architectures requires further analysis. The particular implementation of an initiator accumulation model discussed in Section 3.1 made the strong assumption that the transcription rate is proportional to cell volume (17). It would be interesting to study more detailed network architectures, where this would be a result rather than a model assumption. Models at the molecular level may also begin to investigate the problem of protein number regulation. Since proteins are made by ribosomes, an important question is how ribosomes are allocated toward translating different types of proteins. Quantitative patterns at the bulk level have emerged regarding ribosome allocation (36, 46), but the picture at the dynamical, cell-cycle level is less clear. Models of stochastic gene expression that extend existing ones, which often consider a fixed cell volume (40), to incorporate cell cycle regulation could shed light on this aspect.

Incorporating cell cycle regulation can, in turn, help researchers understand cell size regulation in organisms with a circadian clock, such as mammalian cells and cyanobacteria. Recent works have begun to examine cell size regulation in these organisms (35, 37, 43, 61), and some have suggested that the circadian clock may affect organisms' cell size regulation. How to model these processes at molecular, coarse-grained, and population levels remains an intriguing question.

Since the various classes of models of cell size regulation reviewed above are quite fundamentally different, we conclude by discussing the question of which model to use to analyze what data or to elucidate what phenomenon. First, as discussed in Section 2.4, CRMs differ from DSMs in that CRMs take as a parameter an entire function that describes the instantaneous division rate, whereas DSMs take, for example, only the strength of regulation and the magnitude of the coarse-grained stochasticity (although the form of the stochasticity must be assumed). Moreover, CRMs assume a priori whether regulation depends only on the current size or also on the size at birth, whereas DSMs can be used to determine the mode of regulation. These could be reasons for the increasing visibility of DSMs as models of cell size regulation.

Another fundamental distinction is the difference between division-centric models and models that place control at an upstream event. Division-centric models, such as those described by Equation 9, make the strong assumption that all information relevant for determining division timing is stored in the current size and the size at birth. This does not have to be the case. For example, it is widely accepted that, in *S. cerevisiae*, control occurs over the Start transition (the transition from the G1 phase to the S phase of the budding yeast cell cycle, approximately coinciding with the budding event), and the duration of the budded phase is uncorrelated with size (47). As reviewed briefly in Section 3.1, Equation 9 can be adapted to place control at various cell cycle events, which may then lead to additional predictions that can illuminate the coupling between different cell cycle events.

The above distinction does not imply that we should always use the most detailed description. To cite Levins (31, p. 430], "All models leave out a lot and are in that sense false, incomplete, inadequate. The validation of a model is not that it is 'true' but that it generates good testable hypotheses relevant to important problems." It is often helpful to sacrifice details not pertinent to the phenomenon under consideration. For example, the initiator accumulation model can be considered to trigger division rather than initiation of DNA replication. Yet the simplified model may still provide mechanistic insights into the statistical properties, as discussed in Section 3.1. These mechanistic models are altogether different from the phenomenological DSMs and CRMs.

In summary, the appropriate model to use depends not only on the organism or system in question, but also on the phenomena explored within the model. We have sketched in this review a map of the existing models. Technological advances now enable collection of more accurate and

larger data sets. These will likely stimulate further development of models, which, in turn, will influence experimental directions.

SUMMARY POINTS

1. Studying quantitative patterns associated with cell size and modeling them using stochastic models can shed light on the underlying biological mechanisms.
2. Cell size in *E. coli* can be described by a first-order AR model in which the present value depends only on the value in the previous generation.
3. The initiator accumulation model is a molecular network architecture of cell size regulation that appears to be consistent with existing experimental results in bacteria.
4. Cell size regulation, as long as it is present, does not affect the population growth rate within existing models.

FUTURE ISSUES

1. What is the reason for the prevalence of adder correlations for cell size regulation?
2. What are molecular implementations that regulate the cell cycle in changing environments, and what limits of biological stochasticity can these implementations sustain?
3. How are the copy number of proteins and cell size simultaneously regulated, and how can the resulting statistics be described?
4. What are the relationships of cell size and cell cycle regulation with other cellular processes such as circadian clocks, and what models should we use to describe these relationships?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that may be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

We thank Lingchong You for the permission to use the data from Reference 53. P.-Y.H. was supported by the Harvard Materials Research Science and Engineering Center program of the National Science Foundation under award number DMR 14-20570. J.L. was supported by the Carrier Fellowship. A.A. thanks the Alfred P. Sloan Foundation Research Fellowship, the Harvard Dean's Competitive Fund for Promising Scholarship, the Milton Fund, the Kavli Foundation, and the Volkswagen Foundation for their support.

LITERATURE CITED

1. Adiciptaningrum A, Osella M, Moolman M, Lagomarsino M, Tans S. 2015. Stochasticity and homeostasis in the *E. coli* replication and division cycle. *Sci. Rep.* 5:18261

2. Aldridge B, Fernandez-Suarez M, Heller D, Ambravaneswaran V, Irimia D, et al. 2012. Asymmetry and aging of mycobacterial cells lead to variable growth and antibiotic susceptibility. *Science* 335:100–4
3. Amir A. 2014. Cell size regulation in bacteria. *Phys. Rev. Lett.* 112:208102
4. Amir A. 2017. Is cell size a spandrel? *eLife* 6:e22186
5. Banerjee S, Lo K, Daddysman M, Selewa A, Kuntz T, et al. 2017. Biphasic growth dynamics control cell division in *Caulobacter crescentus*. *Nat. Microbiol.* 2:17116
6. Barber F, Ho P, Murray A, Amir A. 2017. Details matter: noise and model structure set the relationship between cell size and cell cycle timing. *Front. Cell Dev. Biol.* 5:92
7. Box G, Hunter J, Hunter W. 2005. *Statistics for Experimenters*. Hoboken, NJ: Wiley
8. Box G, Jenkins G, Reinsel G. 1994. *Time Series Analysis: Forecasting and Control*. Upper Saddle River, NJ: Prentice Hall
9. Brenner N, Braun E, Yoney A, Susman L, Rotella J, Salman H. 2015. Single-cell protein dynamics reproduce universal fluctuations in cell populations. *Eur. Phys. J. E* 38:102
10. Brenner N, Newman C, Osmanovic D, Rabin Y, Salman H, Stein D. 2015. Universal protein distributions in a model of cell growth and division. *Phys. Rev. E* 92:042713
11. Campos M, Surotsev I, Kato S, Paintdakhi A, Beltran B, et al. 2014. A constant size extension drives bacterial cell size homeostasis. *Cell* 159:1433–46
12. Cooper S, Helmstetter C. 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* 31:519–40
13. Donachie W. 1968. Relationship between cell size and time of initiation of DNA replication. *Nature* 219:1077–79
14. Eun Y, Ho P, Kim M, Renner L, LaRussa S, et al. 2018. Archaeal cells share common size control with bacteria despite noisier growth and division. *Nat. Microbiol.* 3:148–54
15. Fantes P. 1977. Control of cell size and cycle time in *Schizosaccharomyces pombe*. *J. Cell Sci.* 24:51–67
16. Gardiner C. 2009. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Berlin: Springer
17. Ghusinga K, Vargas-Garcia C, Singh A. 2016. A mechanistic stochastic framework for regulating bacterial cell division. *Sci. Rep.* 6:30229
18. Giometto A, Altermatt F, Carrara F, Maritan A, Rinaldo A. 2012. Scaling body size fluctuations. *PNAS* 110:4646–50
19. Godin M, Delgado F, Son S, Grover W, Bryan A, et al. 2010. Using buoyant mass to measure the growth of single cells. *Nat. Methods* 7:387–90
20. Grilli J, Osella M, Kennard A, Lagomarsino M. 2017. Relevant parameters in models of cell division control. *Phys. Rev. E* 95:032411
21. Harris L, Theriot J. 2016. Relative rates of surface and volume synthesis set bacterial cell size. *Cell* 165:1479–92
22. Hashimoto M, Nozoe T, Nakaoka H, Okura R, Akiyoshi S, et al. 2016. Noise-driven growth rate gain in clonal cellular populations. *PNAS* 113:3251–56
23. Hill N, Kadoya R, Chattoraj D, Levin P. 2012. Cell size and the initiation of DNA replication in bacteria. *PLOS Genet.* 8:e1002549
24. Ho P, Amir A. 2015. Simultaneous regulation of cell size and chromosome replication in bacteria. *Front. Microbiol.* 6:662
25. Iyer-Biswas S, Crooks G, Scherer N, Dinner A. 2014. Universality in stochastic exponential growth. *Phys. Rev. Lett.* 113:028101
26. Iyer-Biswas S, Gudjonson H, Wright C, Riebling J, Dawson E, et al. 2016. Bridging the time scales of single-cell and population dynamics. arXiv:1611.05149 [q-bio.QM]
27. Kennard A, Osella M, Javer A, Grilli J, Nghe P, et al. 2016. Individuality and universality in the growth-division laws of single *E. coli* cells. *Phys. Rev. E* 93:012408
28. Kessler D, Burov S. 2017. Effective potential for cellular size control. arXiv:1701.01725 [cond-mat.soft]
29. Koch A. 2001. *Bacterial Growth and Form*. Berlin: Springer
30. Koppes L, Meyer M, Oonk H, de Jong M, Nanninga N. 1980. Correlation between size and age at different events in the cell division cycle of *Escherichia coli*. *J. Bacteriol.* 143:1241–52
31. Levins R. 1966. The strategy of model building in population biology. *Am. Sci.* 54:421–31

32. Lin J, Amir A. 2017. The effects of stochasticity at the single-cell level and cell size control on the population growth. *Cell Syst*. 5:358–67.e4
33. Logsdon M, Ho P, Papavinasasundaram K, Cokol M, Richardson K, et al. 2018. A parallel adder coordinates mycobacterial cell-cycle progression and cell-size homeostasis in the context of asymmetric growth and organization. *Curr. Biol.* 27(21):3367–74
34. Marantan A, Amir A. 2016. Stochastic modeling of cell growth with symmetric or asymmetric division. *Phys. Rev. E* 94:012405
35. Martins B, Tooke A, Thomas P, Locke J. 2017. Cell size control driven by the circadian clock and environment in cyanobacteria. bioRxiv:183558
36. Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, Barkai N. 2017. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife* 6:e28034
37. Mosheiff N, Martins B, Pearl-Mizrahi S, Gruenberger A, Helfrich S, et al. 2017. Correlations of single-cell division times with and without periodic forcing. arXiv:1710.00349 [q-bio.CB]
38. Osella M, Nugent E, Lagomarsino M. 2014. Concerted control of *Escherichia coli* cell division. *PNAS* 111:3431–35
39. Osella M, Tans S, Lagomarsino M. 2017. Step by step, cell by cell: quantification of the bacterial cell cycle. *Trends. Microbiol.* 25:250–56
40. Paulsson J. 2005. Models of stochastic gene expression. *Phys. Life Rev.* 2:157–75
41. Powell E. 1956. Growth rate and generation time of bacteria, with special reference to continuous culture. *J. Gen. Microbiol.* 15:492–511
42. Priestley M. 1981. *Spectral Analysis and Time Series*. Cambridge, MA: Academic
43. Sandler O, Pearl-Mizrahi S, Weiss N, Agam O, Simon I, Balaban N. 2015. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature* 519:468–71
44. Sauls J, Li D, Jun S. 2016. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Curr. Opin. Cell Biol.* 38:38–44
45. Schaechter M, Maaloe O, Kjeldgaard N. 1958. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J. Gen. Microbiol*. 19:592–606
46. Scott M, Gunderson C, Mateescu E, Zhang Z, Hwa T. 2010. Interdependence of cell growth and gene expression: origins and consequences. *Science* 330:1099–102
47. Soifer I, Robert L, Amir A. 2016. Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy. *Curr. Biol.* 26:356–61
48. Sompayrac L, Maaloe O. 1973. Autorepressor model for control of DNA replication. *Nat. New Biol.* 241:133–35
49. Stewart E, Madden R, Paul G, Taddei F. 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLOS Biol*. 3:e45
50. Susman L, Kohram M, Vashistha H, Nechleba J, Salman H, Brenner N. 2017. Statistical properties and dynamics of phenotype components in individual bacteria. arXiv:1609.05513 [q-bio.CB]
51. Taheri-Araghi S, Bradde S, Sauls J, Hill N, Levin P, et al. 2015. Cell-size control and homeostasis in bacteria. *Curr. Biol.* 25:385–91
52. Taheri-Araghi S, Brown S, Sauls J, McIntosh D, Jun S. 2015. Single-cell physiology. *Annu. Rev. Biophys.* 44:123–42
53. Tanouchi Y, Pai A, Park H, Huang S, Buchler N, You L. 2017. Long-term growth data of *Escherichia coli* at a single-cell level. *Sci. Data* 4:170036
54. Tanouchi Y, Pai A, Park H, Huang S, Stamatov R, et al. 2015. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature* 523:357–60
55. Trueba F, Neijssel O, Woldringh C. 1982. Generality of the growth kinetics of the average individual cell in different bacterial populations. *J. Bacteriol.* 150:1048–55
56. Voorn W, Koppes L. 1997. Skew or third moment of bacterial generation times. *Arch. Microbiol.* 169:43–51
57. Voorn W, Koppes L, Grover N. 1993. Mathematics of cell division in *Escherichia coli*: comparison between sloppy-size and incremental-size kinetics. *Curr. Top. Mol. Gen.* 1:187–94
58. Wallden M, Fange D, Lundius E, Baltekin O, Elf J. 2016. The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* 166:729–39

59. Wang P, Robert L, Pelletier J, Dang W, Taddei F, et al. 2010. Robust growth of *Escherichia coli*. *Curr. Biol*. 20:1099–103
60. Willis L, Huang K. 2017. Sizing up the bacterial cell cycle. *Nat. Rev. Microbiol.* 15:606–20
61. Yu F, Willis L, Chau R, Zambon A, Horowitz M, et al. 2017. Long-term microfluidic tracking of coccoid cyanobacterial cells reveals robust control of division timing. *BMC Biol.* 15:11
62. Zheng H, Ho P, Jiang M, Tang B, Liu W, et al. 2016. Interrogating the *Escherichia coli* cell cycle by cell dimension perturbations. *PNAS* 113:15000–5