Nonequilibrium Statistical Mechanics of Dividing Cell Populations

Naama Brenner[1] and Yair Shokef[2]

[1]Department of Chemical Engineering, Technion-Israel Institute of Technology, Haifa 32000, Israel
[2]Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

(Received 25 April 2007; published 27 September 2007)

We present and study a model for the nonequilibrium statistical mechanics of protein distributions in a proliferating cell population. Our model describes how the total protein variation is shaped by two processes: variation in protein production internal to the cells and variation in division and inheritance at the population level. It enables us to assess the contribution of each of these components separately. We find that, even if production is deterministic, cell division can generate a large variation in protein distribution. In this limit we solve exactly a special case and draw an analogy between protein distribution along cell generations and stress distribution in layers of granular material. At the other limit of extremely noisy protein production, we find that the population structure restrains variation and that the details of division do not affect the tail of the distribution.

DOI: 10.1103/PhysRevLett.99.138102 PACS numbers: 87.10.+e, 87.15.Aa, 87.15.Ya, 87.23.Cc

A population of proliferating cells is a stochastic dynamical system far from equilibrium. Proteins and other molecules are constantly produced and degraded; at the same time, cells grow, divide, and inherit their properties to the next generation. These processes are both inherently stochastic, and therefore a population is diverse in its properties, even if genetically homogeneous. The study of protein variation in cell populations has a long history and has recently received renewed experimental and theoretical interest (reviewed by [1–5]). A general feature that emerges from recent experiments is that, even under well-controlled conditions of protein production, steady-state distributions in the population generally exhibit exponential tails and large coefficient of variation (standard-deviation-to-mean ratio) [6,7]. In some cases, this exponential tail can be traced to specific intracellular noise which is exponentially distributed [7–9]. More generally, steady-state distributions can arise from different combinations of intracellular noise and population dynamics [6]. It is therefore of interest to provide a theoretical framework for understanding the contributions of these two processes.

Inspired by recent experimental results, we here present and analyze a minimal model describing the two forces shaping the population protein distribution—intracellular production noise and proliferation dynamics. We define two limiting regimes of parameters in which each of these forces is dominant and compute the steady-state protein distribution in the population in each regime. In the regime where protein production is well-regulated and variation is primarily induced by fluctuations in transmission along generations, we find a nonexponential, nonuniversal behavior. In this regime, our model is analogous to the q model for stress distribution in granular layers, where a universal exponentially tailed distribution was found [10]. The differences between the two models highlight the properties that are special to the structure of a proliferating cell population. In the other regime, where protein production is highly variable, we find that exponential tails arise in several different models and without sensitivity to the division variation. Here the population dynamics act to restrain the total steady-state variation relative to the distribution of protein production in each cell.

We imagine a population as an ensemble of cells driven by two forces: protein production, internal to the cells, and dissipation by cell division and degradation [6]. For long-lived proteins, such as many forms of *gfp* (green florescent proteins) used in experiments [6,11], division is dominant over protein degradation. Making use of the separation of time scales between protein production (normally continuous through the cell cycle) and cell division (occurring on a short fraction of the cell cycle), we express the time evolution of protein content in the cell as a discrete mapping between consecutive generations:

$$x_{n+1} = \mathcal{M}x_n = q_n(x_n + \lambda_n), \tag{1}$$

where x_n is the protein content in a cell immediately following division in generation n. The production force λ_n represents the total amount of protein produced and accumulated in the cell during generation n up to division time; it is a random variable drawn at each generation from a time-independent distribution $\xi(\lambda)$. The division force q_n is the stochastic fraction of protein inherited at division, similarly drawn from $\eta(q)$ [12]. At this stage, we assume that the entire population divides synchronously and write the Liouville equation for the time evolution of the protein distribution at generation n, $P_n(x)$, in these discrete time steps:

$$P_{n+1}(x) = \int_0^1 dq\,\eta(q) \int_0^\infty d\lambda\,\xi(\lambda) \times \int_0^\infty dx' P_n(x')\delta[\mathcal{M}(x') - x]. \tag{2}$$

The nature of cell division imposes the symmetry constraint $\eta(q) = \eta(1-q)$; in particular, $\langle q \rangle = 1/2$. The

model admits a nonzero steady state for every well-behaved distribution of q and λ. Balance between the source and dissipation of this nonequilibrium system is robustly ensured by the fact that dissipation is proportional to the amount of protein in the cell, whereas production is not. At steady state, the dependence on n drops, and, in terms of generating functions, the Liouville equation reads

$$G(s) \equiv \int_0^\infty dx P(x) \exp(-sx) = \int_0^1 dq\, \eta(q) G(qs) H(qs), \tag{3}$$

where $H(s)$ is accordingly the generating function for λ.

If q and λ admit finite moments, then the steady-state distribution of x has the same property: Independence between q and λ enables raising Eq. (1) to the kth power so that averaging over the population yields

$$\langle x^k \rangle = \langle q^k \rangle \langle (x+\lambda)^k \rangle = \langle q^k \rangle \sum_{j=0}^{k} \binom{k}{j} \langle x^{k-j} \rangle \langle \lambda^j \rangle, \tag{4}$$

from which all moments of x may be extracted. In particular, the mean protein level $\langle x \rangle = \langle \lambda \rangle$.

We now consider the limiting case where protein production is perfectly regulated: $\lambda = \text{const}$. In this regime, variation comes from population dynamics, namely, noise in cell division and inheritance. The coefficient of variation then follows from Eq. (4) to be $\sigma_x/\langle x \rangle = 2\sigma_q/\sqrt{3/4 - \sigma_q^2}$, where σ_q is the standard deviation of the division ratio q. Since $0 \le q \le 1$ and $\langle q \rangle = \frac{1}{2}$, $0 \le \sigma_q \le \frac{1}{2}$, and hence $0 \le \sigma_x/\langle x \rangle \le \sqrt{2} \approx 1.4$. If protein is distributed uniformly between mother and daughter cells, $\sigma_q = 1/\sqrt{12} \approx 0.3$ and $\sigma_x/\langle x \rangle = 1/\sqrt{2} \approx 0.7$.

For most proteins, the distribution between mother and daughter cells at division is in correlation with cell size with additional noise. For symmetrically dividing cells, $\eta(q)$ will be centered around $\frac{1}{2}$, whereas for asymmetric division, it will typically have two peaks around the fractions $f \neq \frac{1}{2}$ and $1 - f$. Small noise around symmetric division will result in a narrow limiting distribution, whereas the same is not true for asymmetric division. More quantitatively, with a distribution of width a around the typical division value f, the limiting coefficient of variation $\sigma_x/\langle x \rangle$ will tend to zero as $a \to 0$ for symmetric division ($f = \frac{1}{2}$), while it will tend to a nonzero value $\sigma_x/\langle x \rangle = 2\sqrt{(1/4 - f + f^2)/(1/2 + f - f^2)}$ for the asymmetric case (see Fig. 1). For $f = 0.6$, which is typical for budding yeast cells, we obtain $\sigma_x/\langle x \rangle \approx 0.23$. This is consistent with measurements under conditions where cell size effects were dominant in creating population variation [13]. These experiments were previously explained using a more elaborate model taking into account the specific features of the yeast cell cycle [13].

For a uniform distribution of protein between mother and daughter cells $\eta(q) = 1$, the model steady-state distribution is exactly solvable. Equation (3) then reads

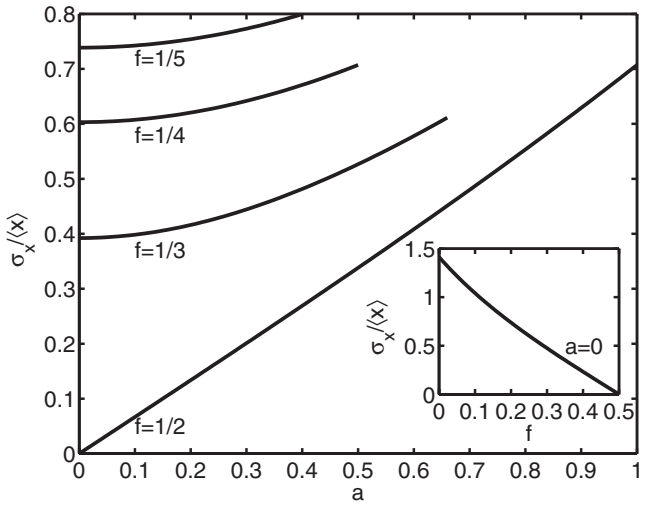

FIG. 1. Coefficient of variation in protein distribution vs division parameters. a represents the variation at cell division, and f is the underlying asymmetry of this division, with $f = 1/2$ corresponding to symmetric division.

$$G(s) = \int_0^1 dq\, G(qs) \exp(-sq\lambda). \tag{5}$$

This, by differentiation, is equivalent to the differential equation

$$G(s) + s\frac{dG}{ds} = G(s) \exp(-\lambda s), \tag{6}$$

which has the exact solution

$$G(s) = \exp\left[\sum_{k=1}^{\infty} \frac{(-\lambda s)^k}{k!k}\right] \equiv \exp[\psi(s)]. \tag{7}$$

This solution provides us with an exact expression for the cumulants of the protein distribution $P(x)$:

$$\kappa_n = (-1)^n \frac{\partial^n \ln G}{\partial s^n}\bigg|_{s=0} = \frac{\lambda^n}{n}. \tag{8}$$

Choosing units of x so that $\lambda = 1$, Fig. 2(a) shows the generating function $G(s)$, decaying as $1/s$ for $s \to \infty$. The probability density $P(x)$ is found by the inverse Laplace transform of the generating function $G(s)$. The asymptotic behavior of $P(x)$ for large x can be computed by the approximation of steepest descent [14]. A single point s^* on the negative real axis, satisfying $x = [1 - e^{-s^*}]/s^*$, gives the leading contribution in this approximation; an infinite number of points with imaginary parts add up incoherently to a negligible contribution. The path of steepest descent is found to be perpendicular to the real axis, finally giving

$$P(x) \sim \frac{\exp[xs^* + \psi(s^*)]}{\sqrt{2\pi}|\psi''(s^*)|^{1/2}}. \tag{9}$$

We note that this is an example of a solvable "inverse problem," where the tail of the probability density $P(x)$ is

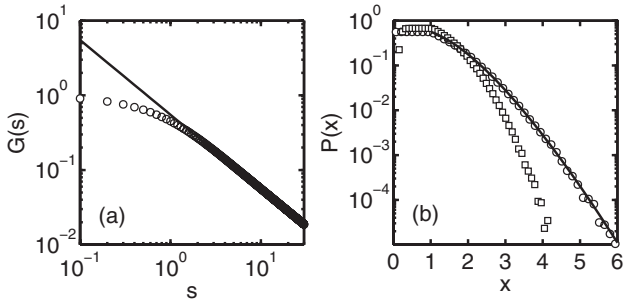

FIG. 2. Population distribution with constant protein production. (a) Generating function $G(s)$ for a uniform division distribution $\eta(q) = 1$, computed exactly as a sum over cumulants (○), with asymptotic approximation $G(s) \sim c/x$ (line; $c =$ const). (b) Monte Carlo simulation of the distribution function $P(x)$ (○), compared to the asymptotic approximation equation (9) (line). Also depicted (□) is the distribution for a division fraction with a smaller standard deviation (square of width $a = 0.8$ around $f = \frac{1}{2}$).

found from the inverse transform of the cumulant-generating function $\psi(s)$ by asymptotic expansion.

Figure 2(b) displays $P(x)$ from Monte Carlo simulations (circles) with the asymptotic approximation Eq. (9) (line), showing excellent agreement. The leading order behavior of the asymptotic approximation is $P(x) \sim 1/x^x$. Also shown in Fig. 2(b) is the Monte Carlo result for a different, narrower distribution of protein at division $\eta(q)$ (squares). The tails of the two distributions cannot be scaled to the same shape; thus, there is no simple universality in the regime of dominant division noise, but rather there is sensitivity to the details of the division function.

It is instructive at this point to draw an analogy between cell populations and other nonequilibrium physical systems. Our model is similar in spirit to stochastic scalar models characterizing fluctuations of stress in static granular packings [10,15] and of energy in dynamic granular gases [16,17]. In the q model for static granular packings [10], grains are assumed to be layered; each grain is characterized by the weight it bears, comprised of the grain's constant self-weight together with the load applied on it by other grains from the layer above it. Each grain randomly partitions its weight among the grains on which it rests. This is analogous to our population model in the regime of negligible intracellular noise, where each cell has a protein content composed of the constant amount it produced (λ) and what it inherited from the previous generation. The protein content is randomly partitioned upon cell division, and divisions occur synchronously in the population, in analogy to the discrete layers of the granular packing. Thus, in both cases, the limiting distribution reflects fluctuations in the transmission along generations. However, the character of the limiting distributions in the two models is very different. While Coppersmith *et al.* [10] found a universal exponential tail to the distributions, we found here a behavior that depends on the division distribution and for the exactly solvable case $\eta(q) = 1$ behaves

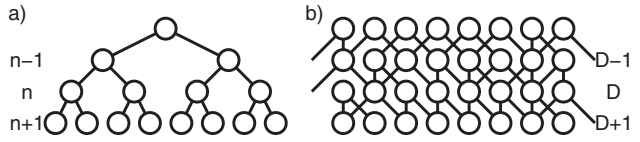

FIG. 3. (a) Topology of the inheritance in a population of dividing cells compared to (b) the q model for stress transmission in granular packings.

as $1/x^x$. One may understand this difference by noting that the transfer between generations (layers) in the two models is differently structured (Fig. 3). In the population model, each cell in generation n has a single ancestor in generation $n-1$ passing proteins to it, whereas in granular packings each grain in layer D carries the weight of several grains from layer $D-1$. This topological difference enhances mixing in the q model and induces a universality in the limiting distribution which is not found in our population model. In practice, while the analytic solution is different, the tails of an exponent and $1/x^x = e^{-x\ln x}$ may be difficult to distinguish in experiment.

We now turn back to our population model in the regime where intracellular protein production is variable. One way to model this regime is to draw λ in Eq. (1) from a broad distribution; specifically, experiments indicate that in some cases this distribution has an exponential tail [7]. Instead, we consider here an exactly solvable model where intracellular variation is induced by an exponentially varying cell lifetime and show below that the steady-state distribution is very similar in the two models.

Consider a population of cells producing protein at a rate μ and dividing with a probability δ per unit time. The protein content $x(t)$ thus evolves according to

$$x(t+\Delta t) = \begin{cases} \underline{\text{Value:}} & \underline{\text{Probability:}} \\ qx(t) & \delta\Delta t \\ x(t) + \mu\delta t & 1 - \delta\Delta t, \end{cases} \tag{10}$$

with q the division fraction as before. The coefficient of variation at steady state is $\sigma_x/\langle x\rangle = \sqrt{(1+4\sigma_q^2)/(3-4\sigma_q^2)} \le 1$; note that the total protein produced per generation is drawn from an exponential distribution with $\sigma_\lambda/\langle\lambda\rangle = 1$. Taking the exponential of Eq. (10) and averaging, one has

$$G(sq) = (1+\gamma s)G(s), \tag{11}$$

with $\gamma \equiv \mu/\delta$ now setting the scale of x, determining its mean value in the population. For $\eta(q) = 1$,

$$(1+\gamma s)\frac{dG}{ds} + 2\gamma G(s) = 0, \tag{12}$$

with the solution $G(s) = (1+\gamma s)^{-2}$. This is readily inverted to yield the Gamma distribution $P(x) = (x/\gamma^2) \times \exp(-x/\gamma)$.

Figure 4 depicts this Gamma distribution, derived above for the uniform division case (line), with Monte Carlo

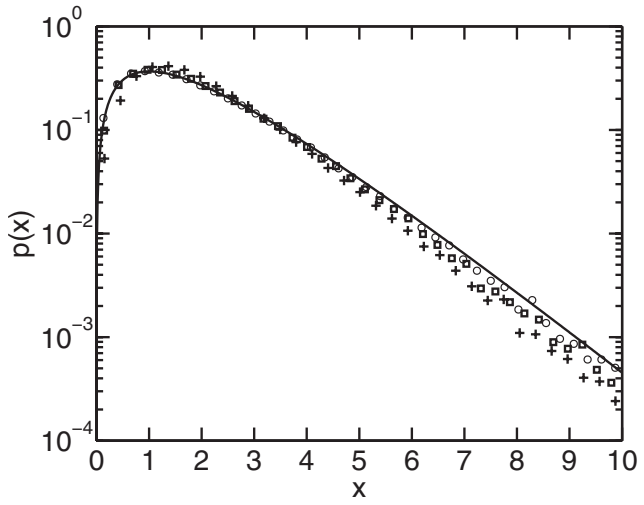

FIG. 4. Protein distributions for highly variable production: the analytical solution (line) and Monte Carlo simulation (○) for a uniform division function; the narrower division function (□, square of width $a = 0.8$ around $f = \frac{1}{2}$, same as in Fig. 2); and the model with synchronous uniform division and an exponentially distributed production (+).

results for different division functions, all displaying similar exponential tails. Also shown for comparison is the distribution for a synchronously dividing population [Eq. (1)], with λ_n drawn from an exponential distribution (+). The behavior of Fig. 4 is also similar to that found in another model with exponential internal protein production [7–9]. We are led to conclude that, in the limit where internal protein production is noisy (here exponential), the steady-state population distribution reflects the internal noise and becomes insensitive to the division details. Moreover, it is also insensitive to the model details.

In all cases mentioned above, the variation in a dividing population, where each cell draws production from a broad distribution, is smaller than the variation in an ensemble of independent particles each drawing its protein content from that same distribution. Thus, the effect of the population is to narrow the distribution from which protein production is drawn. Intuitively, the reason is that roughly half of the protein content is a remnant of the previous generation, and this intergeneration memory reduces the steady-state variation.

In summary, we have presented and analyzed a theoretical model for the nonequilibrium statistical mechanics of a phenotypic quantity (such as protein content) in a dividing cell population. The model enables one to separate contributions to the variation in the population arising from the intracellular noise in production and from variation in the process of division and inheritance. We used the model to calculate the steady-state protein distributions and found that the structure of the biological population affects the limiting distribution in both regimes where either of the processes dominates. For dominant division variation, in an exactly solvable case we found $e^{-x \ln x}$ tails, i.e., nonexponentially decreasing but with a weak asymptotic correction to an exponent. The behavior was found to be sensitive to the details of the division function [10]. In the opposite extreme of dominant intracellular variation, we found that the population structure decreases the variability relative to the bare distribution of protein production. Similar exponential tails can arise either by internal protein production noise or by variation in cell lifetime.

We thank Erez Braun, Shmuel Fishman, Dov Levine, and Yael Roichman for helpful discussions and the Aspen Center for Physics where part of this work was performed. This work was supported in part by the U.S.–Israel Binational Science Foundation (N. B.), by GIF (N. B.), and by the NSF MRSEC program under No. DMR 05-20020 (Y. S.).

[1] H. H. McAdams and A. Arkin, Trends Genet. **15**, 65 (1999).
[2] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Genet. **6**, 451 (2005).
[3] J. M. Raser and E. O'Shea, Science **309**, 2010 (2005).
[4] S. V. Avery, Nat. Rev. Microbiol. **4**, 577 (2006).
[5] D. Longo and J. Hasty, Mol. Sys. Biol. **2**, 64 (2006).
[6] N. Brenner, K. Farkash, and E. Braun, Phys. Biol. **3**, 172 (2006).
[7] N. Friedman, L. Cai, and X. S. Xie, Phys. Rev. Lett. **97**, 168302 (2006).
[8] O. G. Berg, J. Theor. Biol. **71**, 587 (1978).
[9] J. Paulsson and M. Ehrenberg, Phys. Rev. Lett. **84**, 5447 (2000).
[10] S. N. Coppersmith *et al.*, Phys. Rev. E **53**, 4673 (1996).
[11] E. Braun and N. Brenner, Phys. Biol. **1**, 67 (2004).
[12] We assume that the amount of protein in each cell is large and, hence, treat it as a continuous variable. The division fraction q is thus independent of the amount of protein x in the cell. A production force λ which depends on x can describe feedback, i.e., self-regulation. This case will not be discussed here and has some subtle convergence issues.
[13] D. Volfson *et al.*, Nature (London) **439**, 861 (2006).
[14] C. M. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers* (Springer, New York, 1999).
[15] C.-h. Liu *et al.*, Science **269**, 513 (1995).
[16] Y. Srebro and D. Levine, Phys. Rev. Lett. **93**, 240601 (2004).
[17] Y. Shokef and D. Levine, Phys. Rev. E **74**, 051111 (2006).