

# UCSF

## UC San Francisco Previously Published Works

### Title

How Do Cells Adapt? Stories Told in Landscapes

### Permalink

<https://escholarship.org/uc/item/86f866j4>

### Journal

Annual Review of Chemical and Biomolecular Engineering, 11(1)

### ISSN

1947-5438

### Authors

Agozzino, Luca  
Balázs, Gábor  
Wang, Jin  
et al.

### Publication Date

2020-06-08

### DOI

10.1146/annurev-chembioeng-011720-103410

Peer reviewed



# HHS Public Access

Author manuscript

*Annu Rev Chem Biomol Eng.* Author manuscript; available in PMC 2020 July 07.

Published in final edited form as:

*Annu Rev Chem Biomol Eng.* 2020 June 07; 11: 155–182. doi:10.1146/annurev-chembioeng-011720-103410.

## How Do Cells Adapt? Stories Told in Landscapes

Luca Agozzino<sup>1,2</sup>, Gábor Balázsi<sup>1,3</sup>, Jin Wang<sup>1,2,4</sup>, Ken A. Dill<sup>1,2,4</sup>

<sup>1</sup>The Louis and Beatrice Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794, USA;

<sup>2</sup>Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA

<sup>3</sup>Department of Biomedical Engineering, Stony Brook University, Stony Brook, New York 11794, USA

<sup>4</sup>Department of Chemistry, Stony Brook University, Stony Brook, New York 11790, USA

### Abstract

Cells adapt to changing environments. Perturb a cell and it returns to a point of homeostasis. Perturb a population and it evolves toward a fitness peak. We review quantitative models of the forces of adaptation and their visualizations on landscapes. While some adaptations result from single mutations or few-gene effects, others are more cooperative, more delocalized in the genome, and more universal and physical. For example, homeostasis and evolution depend on protein folding and aggregation, energy and protein production, protein diffusion, molecular motor speeds and efficiencies, and protein expression levels. Models provide a way to learn about the fitness of cells and cell populations by making and testing hypotheses.

### Keywords

landscape; adaptation; homeostasis; evolution; fitness

## 1. INTRODUCTION: CELLULAR DRIVING FORCES

Biological cells are adaptive. Change their environment and they respond. In homeostasis, a cell returns to a stable state after it has been perturbed. In evolution, a population of cells becomes better suited to new circumstances. Cell adaptations are reflections of biological driving forces, often expressible as potential functions and visualizable on mathematical landscapes.

Here, we review recent modeling aimed at learning the adaptation code—that is, how adaptive behaviors are encoded within the cell's biomolecules and networks. This research objective is sometimes called genotype to phenotype ( $G_2P$ ) or genotype to fitness ( $G_2F$ ). But

---

dill@laufercenter.org.

### DISCLOSURE STATEMENT

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

here, we prefer the term adaptation code, for it more broadly encompasses other important factors as wellsprings of adaptive behaviors, beyond just genes and mutations alone (Figure 1):

- It incorporates a cell's environment, not just its genes. Survival of the fittest expresses that evolution aims toward matching an organism to its environment. Fitness and forces cannot be understood without accounting for environmental conditions, such as food and nutrient levels, other competing or cooperating organisms, or stressors such as heat or drugs. Biological change can be driven by a changing environment.
- It incorporates a gene's expression, not just its function. Cell fitness depends not only on a protein's efficacy of action but also on a protein's abundance, which is dictated by messenger RNA (mRNA) levels and gene regulatory networks.
- It incorporates a protein's physics, not just its biology. Cell fitness depends on physical properties—the folding and aggregation health of the proteome (called proteostasis), protein diffusion and transport, and cellular balances of energy. These physical properties are relatively universal across the whole proteome, rather than particular to one protein or another. But like protein biology, protein physics also contributes to adaptive forces of homeostasis and evolution and plays a role in cellular growth laws and stress responses.
- It incorporates a cell's fluctuations, not just its average properties. Inside a single cell, concentrations of a molecule can fluctuate, either because of small internal molecule numbers or because a cell changes environments (oxygen levels, nutrients, temperature, or stressors). A cell population can be highly diverse, even when its individual cells are genetically identical. Fluctuations can sometimes promote the average adaptive direction and sometimes oppose it.

This review emphasizes mechanism-based modeling. The aim is to explain adaptive forces in terms of underlying biomolecular actions and networks. We ask questions of “What if,” not just “What is.” This way of knowing is different from that of sequence comparison studies (1), for example. While some aspects of adaptation can be explained by this gene or that mutation, other adaptations are less pinpoint. Even the simplest traits, such as human height, are correlated with tens of thousands of genes. Sometimes cell fitness depends on a protein's stability or its abundance. A protein's stability can be altered by uncountably many different mutations. A protein's abundance can be changed by many changes in a regulatory network. Metaphorically, there are many ways to change traffic lights in a city that speed up traffic flow, but no one traffic light will tell the story. Mechanistic models can help explain these adaptations.

New insights are coming not only from new modeling but also from new laboratory-scale controlled experiments (2–5). On the one hand, evolution in nature has major sources of unpredictability—environments are ever changing; competitors and cooperators come and go; and genomic complexity, redundancy, and natural fluctuations are present. On the other hand, new laboratory evolution experiments offer learning opportunities because internal and external variables can be held constant (6, 7). New data are coming from vast inventories of

DNA and protein sequences, from high-throughput -omics and gene editing (8–22), from single-cell methods (23), and from controlled artificial gene networks (24–29) that can poke and probe cells or can evolve (30–33) just like natural networks do (6, 7, 34, 35). Whereas yesterday's advances gave fine control over proteins, today's give finer control over networks (36–38).

## 2. THE FORCES OF ADAPTATION AND THEIR VISUALIZATION ON LANDSCAPES

The driving forces of nature can be expressed as principles of minimization of potential functions. They are often illustrated on mathematical landscapes as balls rolling downhill. A ball experiencing gravity tends to a minimum of the gravitational potential energy  $U(\mathbf{x})$  as a function of the ball's spatial position  $\mathbf{x}$ . The force on the ball is given by the local slope,  $\mathbf{f} = -dU(\mathbf{x})/d\mathbf{x}$ . Similarly, molecules and materials tend toward thermal equilibrium states, which are at the minima of free energy,  $G(\mathbf{x})$ , as a function of molecular or material degrees of freedom (DOFs),  $\mathbf{x}$ . The slopes of free-energy functions give the forces acting on molecules or materials. Landscapes can be hierarchical: They can describe single-molecule tendencies toward molecular conformational equilibria or material tendencies toward multiple-molecule equilibria. Similarly, landscapes in biology can express behaviors of single cells or whole populations.

Biology's driving forces—in homeostasis and evolution—can seem different from those of chemistry and physics. Biology's forces seem purpose-like—acting to serve the well-being of the organism. Even so, they, too, can be expressed in terms of forces, potentials, and landscapes at the microscopic or macroscopic levels. Figure 2*a* shows a fitness landscape, which illuminates Darwin's principle of survival of the fittest—that is, the evolutionary tendency of a population toward states of maximum fitness for its environment. What is fitness? It depends. At the macroscopic level, in simple cases, a cell population's fitness may just be its growth rate. At the microscopic level, cellular fitness can be the division rate of a cell. In real-world cases, what evolution is optimizing can be complex and is seldom known, but plausible hypotheses can give useful insights.

Figure 2*b* shows an inversion of the fitness landscape, called the fitness potential (it is the negative logarithm of the fitness), which we prefer here because it retains the downhill convention of the metaphor of a ball rolling. Evolution happens in cell populations because DNA and protein sequences mutate and undergo natural selection, leading to increased fitness, like a ball rolling downhill on a fitness-potential landscape. Homeostasis uses similar terms, and similar math (but different variables), and refers to single cells rather than to populations. How can we model the purpose-like actions of single cells and cell populations in terms of the tendencies of biomolecular processes and networks?

Before describing the math, we summarize a few principles of cellular adaptation landscapes. (*a*) They are often high dimensional; there can be many DOFs. Even the simplest cells, like bacteria, have thousands of different types of proteins, each one of which is hundreds of amino acids long; every amino acid comes from a 20-letter alphabet; and each protein's abundance level is controllable (38, 39). (*b*) Landscapes can be bumpy; rolling

balls can get stuck or slowed down by kinetic traps. Even so, at some level, landscapes also must be relatively smooth, at least along some of the DOFs, or else adaptation would be impossible (40). (c) It is known from directed evolution experiments that evolutionary landscapes are dense with biochemical functions. This means that an enzyme can be switched to catalyzing a different reaction often with only a few mutations (41). (d) Proteins can evolve through a relatively small number of complicated routes along which contingencies can matter (42–44), with bottlenecks requiring permissive mutations that stabilize local structures and maintain correct conformational energy balances (45), and some of these routes may require high-order epistasis (i.e., interactions between three or more mutations) (46). (e) Early stages of evolution can be dominated by single mutations, increasing fitness steeply, whereas later stages can involve multiple competing mutations and epistasis, which can be stochastic and slow to reach but then occasionally fast to evolve further (7). (f) Evolution of a protein to a new function often begins without the loss of the previous function, making promiscuous proteins more evolvable (47). (g) While average tendencies can be toward adapted states, stochasticity and fluctuations can sometimes tend to oppose that direction.

How do we model cell adaptation mechanisms? First, we choose whether the model is microscopic (the cell), macroscopic (the population), or both (multiscale). Then, we define (a) the relevant DOFs, (b) the mathematical function that is being optimized, and (c) a mechanism for the fitness as a function of the DOFs. These are rarely known. So, in the spirit of theoretical physics and evolutionary biology, the learnings from modeling come in the reverse direction: First we hypothesize these premises, and then we see and test what they predict against existing data or new experiments.

### 3. THE THEORY OF ADAPTIVE LANDSCAPES

#### 3.1. Simple Homeostasis: The Basic Idea

We first illustrate the basics of simple homeostasis. What balance of processes can hold the concentration  $x$  of some particular biomolecule constant? Consider an mRNA molecule that encodes a protein. Its concentration  $x$  can be increased by an amount  $\Delta x$  in time  $\Delta t$  through a biochemical network that (a) increases the mRNA synthesis (at rate  $J_{\text{syn}} = k_s$ ) or (b) decreases its degradation (at rate  $J_{\text{deg}} = k_d x$ , proportional to the concentration) (Figure 3):

$$\frac{\Delta x}{\Delta t} = J_{\text{syn}} - J_{\text{deg}} = k_s - k_d x. \quad 1.$$

Homeostasis is defined as the steady state where  $\Delta x / \Delta t = 0$ , that is, where  $J_{\text{syn}} = J_{\text{deg}}$  (see Figure 3). Let us express the tendency to sustain the steady state as a biochemical potential,

$$\phi(x) = \int_{x_0}^x (J_{\text{syn}} - J_{\text{deg}}) dx' = -k_s(x - x_0) + k_d(x^2 - x_0^2)/2. \quad 2.$$

The tendency toward homeostasis can be likened to a ball rolling downhill on this one-dimensional landscape, where

$$\text{force} = \frac{\Delta x}{\Delta t} = - \frac{d\phi}{dx}. \quad 3.$$

In homeostasis, the force (net rate) is the slope of this biochemical landscape, which equals zero at the well bottom. This force is a tendency to restore, resulting from an imbalance of flows. When  $x$  is small, synthesis dominates, increasing  $x$ . When  $x$  is large, degradation dominates, reducing  $x$ .

This example shows how opposing processes of synthesis and degradation can result in homeostasis, a fixed stable concentration. These flows can be expressed as a net tendency toward the minimum of a biochemical potential. Such landscapes can also have multiple minima—that is, multiple homeostatic concentrations (48–52). One main subpopulation of cells would have a stable concentration  $x_1^*$ , and another main subpopulation would have a biomolecule concentration  $x_1^*$ .

Biochemical potential landscapes underlie cellular homeostasis, but they are not fitness landscapes. In constant environments, evolution drives the regulatory networks controlling synthesis and degradation rates  $J_{\text{syn}} = k_s$  and  $J_{\text{deg}} = k_d x$  such that biochemical and fitness-potential minima coincide. Since fluctuating environments imply fluctuating fitness-potential landscapes within different peaks, regulatory networks can also evolve to give rise to multiple biochemical potential minima when environmental conditions fluctuate.

This example is a useful deterministic and continuum approximation when fluctuations are negligible. But in some cases, we also want to account for noise from the few-particle internal fluctuations or from fluctuations of external origin. Internal fluctuations arise from the discrete nature of individual molecules and can be expressed as birth and death events in a master equation for the probability  $P_n$  to have exactly  $n$  molecules in a given cell at time  $t$ :

$$\frac{dP_n}{dt} = k_s P_{n-1} + (n+1)k_d P_{n+1} - (nk_d + k_s)P_n. \quad 4.$$

Below, we generalize the mathematics to handle more complex systems and treat evolution as well as homeostasis. It is easy to see how the math of homeostasis also applies to evolution; in the latter case, the driving mechanisms are mutation and selection instead of synthesis and degradation, but, as we show below, the dynamics of the two processes can be described in a similar way.

### 3.2. General Dynamic Theory of Adaptive Behaviors

A more general way to treat the dynamics of the changing concentrations of genes and proteins inside a single cell—or the changing allele frequencies in an evolving cell population—is through the Fokker–Planck diffusion equation. We start with the ordinary differential equation

$$\frac{dx}{dt} = \mathbf{F}(\mathbf{x}), \quad 5.$$

where  $d\mathbf{x}/dt$  is the rate of change of some concentration (or population or allele prevalence) with respect to time  $t$  (see Figure 4). When describing homeostasis,  $\mathbf{x}$  could be the set of concentrations of different proteins. When describing evolution,  $\mathbf{x}$  is the collection of the frequencies of each allele (i.e., populations of different forms of a gene or protein). Since each species (proteins or genes) can have multiple values of concentrations/frequencies, say  $N$ , a network of  $M$  species will have a large number of states,  $N^M$ .  $\mathbf{F}(\mathbf{x})$  is the vector of driving forces for changing biochemical concentrations. Equation 5 describes forces as velocities or rates (like Newtonian particles in an overdamped regime, where acceleration terms are negligible).

Equation 5 is a fundamental equation for nonlinear dynamics (53). Such models are often studied by identifying the fixed points and performing a linear stability analysis to figure out the stability of each fixed point. However, this approach explores only small local changes (metastability), and not the larger-scale dynamics (global stability) such as the transitions between the fixed points. Moreover, Equation 5 does not account for intrinsic and extrinsic fluctuations. For fluctuations, we need stochastic dynamics, not deterministic dynamics. Fluctuations can be treated either through the master equation approach (Equation 4) or more simply by the Langevin equation approximation (8–11),

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) + \xi(\mathbf{x}, t). \quad 6.$$

The function  $\xi(\mathbf{x}, t)$  is a fluctuating force in time, with a given probability distribution, that adds to the deterministic model of the network dynamics. The statistical properties of the stochastic force are taken to be  $\langle \xi(\mathbf{x}, t) \rangle = 0$  and  $\langle \xi(\mathbf{x}, t) \xi^T(\mathbf{x}, t') \rangle = 2\epsilon \mathbf{D}(\mathbf{x}) \delta(t - t')$ , where  $\epsilon$  is a scale factor quantifying the fluctuation strength and  $\mathbf{D}(\mathbf{x})$  is the diffusion matrix giving rise to the fluctuation correlations (8–11). Since there are many components of the fluctuation sources, it is generally assumed that the stochastic force follows a Gaussian distribution. Figure 4 shows how this formalism can be applied to both cellular homeostasis and evolutionary dynamics.

Now, while Equation 6 describes the dynamics of the mean value and variance of  $\mathbf{x}$ , we often want to know, more generally, the dynamics of the whole probability distribution function  $P(\mathbf{x}, t)$ . This is given by the Fokker–Planck diffusion equation (8–11):

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = -\nabla \cdot \left\{ \mathbf{F}(\mathbf{x}) P(\mathbf{x}, t) - \epsilon \nabla \cdot [\mathbf{D}(\mathbf{x}) P(\mathbf{x}, t)] \right\}. \quad 7.$$

The content of the curly brackets can be defined as a function  $\mathbf{J}(\mathbf{x}, t)$ ,

$$\mathbf{J}(\mathbf{x}, t) = \mathbf{F}(\mathbf{x}) P(\mathbf{x}, t) - \epsilon \nabla \cdot [\mathbf{D}(\mathbf{x}) P(\mathbf{x}, t)], \quad 8.$$

such that Equation 7 takes the form of a continuity equation, expressing the conservation of probability

$$\partial_t P(\mathbf{x}, t) + \nabla \cdot \mathbf{J}(\mathbf{x}, t) = 0; \quad 9.$$

therefore, the function  $\mathbf{J}(\mathbf{x}, t)$  is the probability flux.

Equations 7 and 8 are linear (in  $P$ ) and deterministic. So, while the individual trajectories themselves are not predictable, the dynamics of the statistical distribution is predictable. Furthermore, these expressions give a global description of the system (not limited to local linear stability analysis around fixed points), and this allows us to comprehend all the basins on the landscape and their connectivities.

This formulation illuminates an important principle of forces and flows. It says that there are two types of driving forces (more mathematical details are given in the following paragraphs). The first type is familiar; these forces can be computed from a slope on a potential landscape, as befits the metaphor of a ball rolling downhill. The second type is less familiar; these are forces that are not expressible as a slope on a landscape. The latter is a uniquely dynamical phenomenon. Sometimes, when energy is flowing into a nonequilibrium system, it acts as a force to drive balls to roll around on perfectly level paths on landscapes, where the slope is zero. How do we distinguish between these two types of forces? For any system at steady state, by definition,  $\partial P_{ss}/\partial t = 0$  (8–11), so according to Equation 9,  $\nabla \cdot \mathbf{J}_{ss} = 0$ ; this is called the zero-divergence condition. The two types of forces correspond to the two different ways steady-state systems can achieve the zero-divergence condition. First, zero divergence results when the steady-state flux itself is zero,  $\mathbf{J}_{ss} = 0$ , implying no net flux in or out and implying that the system is at equilibrium and satisfies the principle of detailed balance (8–11). If we imagine rain falling into a well, steady state is achieved only if water is hauled out in buckets at the same rate that rain is filling up the well.

A second way to achieve zero divergence,  $\nabla \cdot \mathbf{J}_{ss} = 0$ , applies when  $\mathbf{J}_{ss} \neq 0$ . In this case,  $\mathbf{J}_{ss}$  has a rotational nature; its force lines typically circulate in loops (8–11, 16, 54, 55). This is called the curl flux. Now, if we imagine that our water well has a horizontal ledge that forms a circular trough located halfway down the well, the rainwater has another option: It can swirl. It pours into the trough, converting vertical rain flow into horizontal flow around and around the ledge. This type of flow, perpendicular to the well axis, never changes the level of water in the well, so it, too, satisfies the requirement that the system be in steady state.

By rearranging the steady-state Equation 8, we find the driving force to have three components (16, 22, 54–56):

$$\mathbf{F} = -\epsilon \mathbf{D} \cdot \nabla U + \mathbf{V}_{ss} + \epsilon \nabla \cdot \mathbf{D}. \quad 10.$$

Here,  $U = -\ln P_{ss}$  can be defined as the potential landscape, and  $\mathbf{V}_{ss} = \mathbf{J}_{ss}/P_{ss}$  is the flux velocity. Thus, the driving force of nonequilibrium dynamical systems can be partitioned into three parts: the gradient-like force of the potential landscape  $U$ , which is associated with the steady-state probability distribution  $P_{ss}$ ; the rotational-like force  $\mathbf{V}_{ss}$ , related to the steady-state probability flux  $\mathbf{J}_{ss}$ ; and the fluctuation-induced force originating from the fluctuations, which vanishes when  $\mathbf{D}$  is independent of  $\mathbf{x}$ . In systems at equilibrium, the net flux vanishes, and  $\mathbf{J}_{ss} = \mathbf{V}_{ss} = 0$ . In such cases, the global minimum of the potential landscape defines the global stability of the system, and the forces are given by the gradient of the landscape. But in nonequilibrium systems, where net flux does not vanish and  $\mathbf{J}_{ss}$  and  $\mathbf{V}_{ss}$  are not zero, the nonequilibrium potential landscape can still be used to quantify the system's global stability together with the flux. The forces and dynamics of the system are

now determined by both the gradient of the potential landscape and the curl force of the flux velocity.

In the zero-fluctuation (i.e., deterministic) limit, the driving force can be decomposed into an intrinsic potential  $U_0 = \lim_{\epsilon \rightarrow 0}(\epsilon U)$  and an intrinsic flux velocity  $\mathbf{V}_0$  (9, 22, 37, 55, 57, 58):

$$\mathbf{F} = -\mathbf{D} \cdot \nabla U_0 + \mathbf{V}_0. \quad 11.$$

The intrinsic potential always decreases along the deterministic trajectory; therefore, it can be used to quantify the global stability of the deterministic nonequilibrium system. The underlying dynamics of the deterministic nonequilibrium system is determined by both the intrinsic potential landscape and the intrinsic flux velocity. The gradient of the intrinsic landscape is perpendicular to the intrinsic flux velocity in the deterministic case. While we describe steady states above, this formulation is readily generalized to handle transient dynamics, time-dependent external conditions, multiple-state-transition mechanisms, and spatially extended systems (22, 55). In the following sections, we discuss the important implications of curl-flux dynamics for homeostasis and evolutionary dynamics.

### 3.3. Static Fitness Landscapes Have Energy-Like and Entropy-Like Components

Like free-energy landscapes in thermodynamics, fitness landscapes have two components: One is energy-like, and one is entropy-like. Imagine one cell behavior  $A$  that can be achieved by  $N_A$  different sequences and another behavior  $B$  that can be achieved by  $N_B$  different sequences. On the one hand, evolution may tend toward  $B$  if those sequences have greater fitness. On the other hand, evolution may also tend toward  $B$  simply if the  $B$  sequences are more numerous. This distinction can be expressed in energy-like and entropy-like terms. The rate at which a protein molecule evolves is given by the dependence on time  $t$  of the probability  $P_i(t)$  that a protein sequence  $i$  becomes fixed in a population by time  $t$ , through mutation and selection. The equilibrium distribution of such probabilities is a Boltzmann-like exponential (59–63):

$$P_i^* = g_i \frac{e^{-\lambda V_i}}{Q}, \quad 12.$$

where  $V_i$  is the fitness potential, a function of the different mutations of a given protein (40), related to the fitness  $f_i$  by  $V_i = -\log f_i$  (59);  $g_i$  is the sequence degeneracy, that is, the number of different sequences of a given fitness;  $\lambda$  is the selective pressure; and  $Q = \sum_i g_i e^{-\lambda V_i}$  is the sum over the relative populations of the different sequences of the protein.

In short, like the equilibrium thermodynamics of materials, evolutionary changes can be energy-like or entropy-like. Energy-like refers to cases where one sequence is preferred to another because of the higher fitness of that specific sequence. Entropy-like refers to cases where one whole category of sequences is preferred to another whole category of sequences simply because the former category has more sequences in it than the latter has. Entropy plays this role: Sometimes a cell population will not converge to a single perfect (maximally fit) amino acid sequence because there are so many alternative sequences that can achieve a sufficient (i.e., near-perfect) fitness instead. Note that these energy-like and entropy-like

components of evolution should not be confused with the thermal energies and entropies of materials. In thermal materials, the balance between energy and entropy is dictated by the temperature; in evolution, the balance of these tendencies is dictated by the selective pressure. Apes can evolve into humans by selective pressures, but not by a change in temperature. Equation 12 provides a general framework to address some questions of protein evolution in the Sections 3.4 and 4.1–4.3.

### 3.4. Evolution Speeds Range from Days to Millions of Years

Evolution can happen over a wide range of timescales. Some evolutionary processes take millions of years; others take weeks to months. What explains this broad range? Modeling shows that if a landscape of fitness potential is flat, evolution is slow; if it is funnel shaped, evolution can be fast. The term selection strength in evolutionary biology corresponds to the slope. Other factors affecting the speed are the rate of mutations and the effective population size (64, 65). To study the principle of how evolution speed depends on landscape shape, Equation 12 is combined with the simplest nonflat landscape, which has a slope that is linear in the number  $m$  of mutations;  $V(m) = \text{constant} \times m$  away from the optimal sequence (63). The dynamics of this model is readily solved analytically in the limit of low mutation rate and large population size to give the adaptation time,  $\tau_{\mathcal{A}}$ , for a protein to reach its optimum sequence:

$$\tau_{\mathcal{A}} \simeq \frac{(1 + ze^{-\lambda V_0})^L}{\omega_0 L}, \quad 13.$$

where  $z$  is the number of possible mutations a residue in the protein can have relative to its starting sequence ( $z = 19$ ),  $L$  is the total number of residues in the protein,  $\omega_0$  is the average fixation rate for a single point mutation, and  $V_0$  is the slope of the fitness potential. What explains the large dynamic range in evolutionary adaptation is that  $\tau_{\mathcal{A}}$  is an exponential of an exponential function in Equation 13 (Figure 5). All of this refers to a constant environment providing a constant landscape; changing environments can renew the driving forces of evolution.

## 4. PROTEIN FOLDING AS AN ACTOR IN HOMEOSTASIS AND EVOLUTION

### 4.1. Changing the Temperature Changes Proteome Folding Stabilities

How do cells adapt when put into an unfamiliar temperature? For simple cells, the fitness can be taken to be the growth rate. And the growth rate dependence on temperature is often known. Simple cells grow fastest at the temperature of their natural environment. Figure 6b shows the thermal growth law of *Escherichia coli*: Its growth rate as a function of its growth temperature has a peak. If a cold cell is heated up, it grows faster. This resembles simple Arrhenius-like chemical kinetics.

But if a cell is heated further, its growth rate slows down sharply. Why? The proposition of the thermal proteome unfolding model is that the plummeting growth rate, and death, when cells are too hot results from the denaturation of the proteome. In this model, the fitness

potential  $V(T)$  as a function of temperature  $T$  is taken to be proportional to the number of proteins that are folded:

$$V(T) = - \frac{A \exp(-\Delta G/RT)}{1 + \exp(-\Delta G/RT)}, \quad 14.$$

where  $A$  is the protein abundance and  $G(T, L) = H(L) + C_p(L)(T - T_h) - T S(L) - T C_p(L) \ln(T/T_s)$  is the free energy of folding as a function of chain length  $L$  in terms of known enthalpy  $H$ , entropy  $S$ , and heat capacity  $C_p$  and measured temperature constants  $T_s$  and  $T_h$  of average proteins (66–68).

Cooled cells (Figure 6a) grow faster when heated. This increasing function can be fit by an Arrhenius temperature law, indicating that growth rates of cooled cells may be governed by one or more key biochemical processes. Warmed cells (Figure 6b) slow their growth dramatically upon heating and die. For many free-growing organisms, the peak point—that is, the temperature of fastest growth—happens to approximately equal the temperature of the cell's natural environment.

What do we learn from the thermal proteome unfolding model? First, it shows how a cell-level phenotype, its heat stress behavior, can be computed from physical properties of the cell's biomolecules—namely, its proteome unfolding behavior. The thermal parameters of proteome unfolding are known from in vitro experiments on different proteins (67). Second, while this model describes the behavior of a given cell under varying temperatures, it also gives insights into how cell populations evolve under pressures to survive at different temperatures. A cell population can evolve to grow rapidly at an unfamiliar temperature via mutations that change the stability profile of the proteome's average protein (see Figure 6).

Interestingly, individual cells may respond differently to temperature changes (69). In a population of heat-shocked yeast cells, a subpopulation was observed to become resistant and continue growing (albeit slower), while another subpopulation stopped growing completely and degraded its own proteins (70). Thus, different subpopulations of cells can respond quite differently to temperature, and possibly to other stresses.

#### 4.2. Cells Acclimate Faster to Hotter Than Colder Environments

This growth law (growth rate versus temperature) is modeled through protein folding stability. Now, this same mechanism can predict how rates of protein evolution depend on temperature. How fast can a bacterial protein evolve when transferred into an environment of different temperature? The prediction below is that bacteria can adapt much faster to a warmer climate than to a colder climate. In Figure 6a, the mesophile that lives at 40°C is at a ridge peak of fitness because freestanding cells grow the fastest at the temperature of their natural environment. Suppose the mesophile is now upshifted to live at 70°C. Path ① indicates that, before any mutations have occurred, this upshift causes the mesophile to grow slower. Path ② indicates that, over time, mutations occur (30 are shown in the figure) that cause the protein to become well adjusted to the warmer environment. Alternatively, a thermophile that is preadapted to 70°C adapts to a colder environment following paths ③ and ④. Adaptation is much faster (up to 5,000-fold) along path ② than along path ④ (63).

What is the mechanism? Folded proteins are destabilized by heating, but not by cooling. A cell with sicker proteins adapts fast (path ②) because it is climbing a steep slope on a fitness landscape. In contrast, cooling does not destabilize folded proteins, so cells are slower to adapt to cooling (path ④). In summary, cells should adapt to warm climates faster than to colder ones (63). This prediction has not been tested, as far as we know, but experiments show that a mesophile can successfully adapt to a warmer environment (71).

#### 4.3. Proteins That Are Least Abundant in a Cell Are Fastest to Evolve

Through a cell's evolution, the expression level of a gene (and thus the abundance level of its protein) can change (Figure 7). Proteins that are abundant tend to evolve slowly (63, 72–80). This is called the expression level–substitution rate anticorrelation. It has a simple explanation. If a mutation diminishes the fitness of a given type of protein, then the more abundant that protein is, the more deleterious that mutation is overall to the cell. If that mutation causes misfolding or reduces folding stability, the cell's fitness  $V$  is reduced by the protein's abundance  $A$  (Equation 14). Or, if that mutation causes protein aggregation (79), then the effect on cell fitness is proportional to  $A^2$  (63). Such protein folding contributions to fitness and evolution rates successfully predict the mutational fitness effects in viruses and simple cells (63, 81).

#### 4.4. Proteostasis Is a Well-Oiled Protein Homeostasis Machine

Here we describe a model of proteostasis, the homeostatic maintenance within a cell of the folded states of cell's proteins (collectively, its proteome). Folded proteins are maintained by flows from protein synthesis and degradation and by an adenosine triphosphate (ATP)-driven cellular machinery of chaperones that tip the cell's kinetics in favor of folded states. This balance in *E. coli* bacteria has been expressed in the hospital model of proteostasis (HMOP) (82). This name signifies how folding sick proteins entails decision-making like that in hospitals that treat sick patients: (a) identify which protein is sick (misfolded or aggregated), and how sick it is, and (b) send it to the right doctor (i.e., the right chaperone system) to fix its folding/aggregation problem.

HMOP is composed of (a) prior biophysical models of spontaneous conformational change (folded, unfolded, misfolded, aggregated) of different classes of proteins and (b) coupled ordinary differential equations describing the trafficking flow of different proteins through different chaperones (see Figure 8). Nodes represent the different states of protein conformation and/or the binding of proteins to chaperones. Edges represent traffic flows. HMOP is a deterministic model, not including fluctuations, that has been applied to steady state to compute the steady-state average concentrations of all the node species and the fluxes between them. This modeling describes how stable proteostasis states are encoded within this particular physicochemical network. And it allows for studies of perturbations of these average states, either those that are imposed on a single cell with given rate coefficients or those that are imposed on cell populations that can evolve different values of rate coefficients.

Figure 8b shows one model prediction: Very sick proteins (stuck in deep misfolded kinetic traps) traffic mostly through the GroEl chaperone system. Several matters of principle are

resolved by the model that had not been obtainable from experiments alone: (a) It shows how complex decisions are encoded in this cell-wide asynchronous physicochemical network to find, identify, sort, and fix sick proteins. (b) It shows that the central physical property of a client protein that defines its trafficking is its dwell time in misfolded states. (c) It shows that this process is energy efficient, as the sickest proteins use the most energy-expensive chaperones. (d) It shows that chaperone levels are adaptive, increasing at very fast growth rates to prevent rapidly produced proteins from aggregating and increasing at very slow growth rates to prevent protein degradation from costing the cell energy when it is not growing.

## 5. SOME ADAPTATIONS ARE IN PROTEIN PHYSICAL PROPERTIES

### 5.1. Salt Growth Laws Depend on Protein Diffusion Rates

Cells grow slowly when put into media with high salt concentrations. This has been modeled as an evolutionary adaptation of the protein density in the cell. Bacterial cells are densely packed with proteins, to a density of about 20% (68). In the protein transport rate model, adding external salt increases the osmotic pressure and draws water out of the cell, densifying the cell's contents and squeezing together the proteins, which causes the proteins to diffuse more slowly due to crowding (68). The evolutionary DOF is taken to be  $c$ , the crowding (i.e., the protein density in the cell). The fitness proposition is that evolutionary changes can alter  $c$  and do so in a way that maximizes the rates of transport of proteins within the cell. If a type of protein is too crowded, it diffuses in the cell slowly; if a type of protein is too dilute, its net flux is small. This is described by the diffusion expression

$$\text{rate} = cD(c) = \frac{kT}{6\pi\eta a} \left[ c \left( 1 - \frac{c}{c_{\text{xtal}}} \right)^2 \right], \quad 15.$$

where  $D$  is the diffusion constant of a protein (assumed to be spherical) in the cell,  $kT$  is Boltzmann's constant times temperature,  $\eta$  is the solvent viscosity,  $a$  is the protein's radius, and  $c_{\text{xtal}}$  is the maximum concentration achievable by sphere packing of the proteins. To predict the maximum fitness, we take the derivative  $d(\text{rate})/dc$  and set it equal to zero. This gives the value  $c = c_{\text{xtal}} = 0.20$ , which is consistent with the observed protein density inside *E. coli*. And Equation 15 gives the functional form of the salt growth law shown in Figure 9. Figure 9 shows additional confirmation of the protein transport rate model, namely, that added salt slows the diffusion of green fluorescent proteins in single-cell experiments (68) by two orders of magnitude.

### 5.2. Energy Efficiency Is a Fitness Function for Protein Production in Fast-Growing Cells

Simple cells grow and duplicate faster in media with more food. Added sugar leads to an upshifted concentration,  $R$ , of ribosomes relative to the concentration,  $P$ , of nonribosomal proteins. Many cell processes are involved in this global readjustment with growth rate. What cell fitness property is being optimized in the evolution of the many microscopic parameters involved? One hypothesis is that the cell simply aims at maximizing its duplication speed. But if so, cells ought to approach the ratio  $\alpha = P/R \rightarrow 0$  in the limit of plenty of food. Instead, it is observed that  $P/R \rightarrow 0.6 - 0.8$  (Figure 10). The propositions of

the ribosomal upswitch model (83, 84) are that the cell's ratio of  $P$  to  $R$  is determined by its concentration of ATP ( $A$ ) and that the fitness function for well-fed cells is energy efficiency, that is, doubling rate per ATP molecule. The steady-state limit of this nonlinear model can be solved as a third-order polynomial for the fitness  $\epsilon(\alpha)$  as a function of the evolutionary DOF  $\alpha$ , as shown in Figure 10. The maximum fitness is computed by setting  $d\epsilon/d\alpha = 0$  and gives a value of 0.72, which is consistent with the data.

### 5.3. Protein Motors and Pumps Can Trade Off Speed Versus Efficiency

Much of a cell's energy is used by its motors, such as  $F_0F_1$  ATPase (see Figure 11), and pumps. It may have been evolutionarily advantageous to optimize speed, efficiency, or some other property. A two-state kinetic model has been developed that describes a broad range of such motors by allowing for evolutionary DOFs, such as where the kinetic barrier steps happen in the motor cycle (85). By fitting these few parameters in simple models to nearly a dozen different biomolecular motors and pumps, we can learn what, if anything, is optimized and might serve as a fitness quantity.

Figure 11 shows that there are indeed properties that appear to be optimized for molecular motors and pumps. In these calculations, it is supposed that biomachines evolve only within a restricted range of operating specifications: (a) Their input is restricted to a small range of chemical free energies since their power is supplied by ATP. (b) Therefore, their output work per cycle cannot exceed this amount either. This model creates the freedom to look at different possible fitness functions and ask which such fitness function—if any—is maximized by the properties of known motors and pumps. Figure 11 shows that the fitness function satisfying this condition is the output power per unit input energy, where power is the work performed per unit time.

As a different mechanical adaptation, Schuech et al. (87) have shown that the shapes of bacteria—from straight to curved rods—can be understood as evolutionary Pareto optima of three properties: shapes that favor good swimming speeds, shapes (more rodlike) that give a better signal-to-noise ratio in detecting chemical gradients (fundamental for chemotaxis), and shapes that reduce the cost of cell construction.

## 6. PROTEIN ABUNDANCES ARE RESPONSIVE TO THE ENVIRONMENT

### 6.1. Protein Concentrations Result from a Balance of Factors

The vignettes above focus on physical properties of proteins—stability, aggregation, diffusion, and motor behaviors. Also important are protein abundances (concentrations) in the cell. Changing a cell's environment can lead to changing concentrations of the proteins inside it. Fitness appears to be a Goldilocks balance. If a protein is too dilute, its effect on cell health is too feeble. If a protein is too abundant, the cell has wasted energy by overproducing it and is evolutionarily uncompetitive (88). Figure 12a shows experimental evidence for this latter contributor to fitness. Perfeito et al. (89) inserted the gene coding for the LacZ protein as an unnecessary gene in *E. coli*. The figure shows that increasing its concentration reduces the cell's fitness. Interestingly, at sufficiently high protein concentrations, there is a sharp drop in fitness. This nonlinearity of fitness is the

consequence of a feedback loop: The production and the activity of Lac proteins affect the growth rate, which, in turn, modulates the density of these molecules.

The costs of excess protein levels can arise not only from protein synthesis but also indirectly through protein function (90, 91). In eukaryotes, the function of transcriptional activator proteins is often costly (92) because they recruit general transcription factors that are needed cell-wide. If such an activator protein is in excess, it will tie down the general transcription factors at particular genes, depleting their presence elsewhere and indirectly lowering the expression of other important genes. This is generally true for cellular resources such as ribosomes, polymerases, and global regulators: If they are tied down somewhere, the cell will suffer due to their depletion (93).

## 6.2. Drugs Can Diminish the Fitness of a Cell, Sometimes Sharply

Bacterial growth slows down when antibiotic drugs are added to the media. Bacterial growth rates reflect the fitness of bacteria for their environment, and bacteria are less fit for environments that contain drugs. Growth speeds diminish in a sigmoidal way as drug concentrations increase (see Figure 12*b*). Now, consider a multidimensional fitness landscape. What happens when we breed drug-resistant cells? Figure 12*b* shows two consequences. First, by definition, we need higher drug concentrations to slow the growth of drug-resistant cells than we need for wild-type cells. Second, in highly resistant cells, the transition to zero growth rate is very sharp, showing a bistability. The basis for this nonlinearity and bistability appears to be a positive feedback mechanism generated by an innate global effect of drug-inhibited growth on gene expression: Translation-inhibiting antibiotics reduce growth and thereby reduce gene expression, including expression of genes conferring drug resistance, increasing the effect of the drug in a positive feedback loop (91, 94). The positive loop is responsible for bistability (49, 95).

## 6.3. Changing Environments Can Be Confusing; Cells Can Still Find the Optimal States

Consider a cell that lives in an environment  $E_1$ . Suppose a gene  $G_1$  encodes a protein  $P_1$  that confers a benefit to the cell in  $E_1$ . Then the cell will evolve its regulatory network to express the optimal level of  $P_1$ . Now, we see if we can confuse the cell. The cell cosynthesizes with  $P_1$  another protein  $P_2$  that confers a cost when the cell is in a different environment  $E_2$ . Now, we cause confusion by driving the environment to fluctuate between  $E_1$  and  $E_2$ . Or we have both genes turn on in  $E_2$  and turn off in  $E_1$ , maximizing the cost of  $P_2$  and minimizing the benefit of  $P_1$ . Can the cell evolve a regulatory network that optimizes the levels of both  $P_1$  and  $P_2$  to handle this fluctuating environment? Figure 12*c* shows that, indeed, the population adapts toward the optimal solution in a genetic module subject to the *lac* repressor in *E. coli* (3, 4). Evolution of the regulatory net proceeds during three cycles of mutation and selection. Evolution is delayed in crossing a fitness moat of confusion, requiring a few rare mutations to reach the high-fitness state of the cell, embodied by a new protein that responds inversely to the external inducer concentration.

## 6.4. Cell Enzymes Should Be Neither Too Dilute nor Too Concentrated

A cell's fitness depends on its enzyme concentrations. Too little enzyme means too little metabolic flux of a particular pathway to serve the needs of the cell. Too much enzyme

means the cell overinvests in producing it, and the cell cannot compete with other cells that are more frugal. Each enzyme should achieve a balance: neither too little nor too much to contribute to a biochemical pathway that is properly balanced with other pathways. What determines how much enzyme is too much? The enzyme concentration should be below the  $K_m$  (Michaelis–Menten binding constant) for the substrate, or else enzyme is wasted (96). In addition, the enzymes' actions, and not just their expression levels, should be under evolutionary pressure. The enzymes of an optimal cell should have high catalytic efficiencies (97).

The levels of enzymes in a pathway depend on the efficiency of subsequent enzymes. This is important to avoid the accumulation of toxic intermediate metabolites that can occur if an enzyme's flux exceeds what the next enzyme can process. For example, excess lactose flux imported into the *E. coli* cell can be toxic (98). To avoid this, cells regulate enzyme levels according to the fluxes through metabolic pathways.

In Sections 6.5 and 6.6, we describe studies in which gene expression levels are controlled through the insertion of synthetic gene circuits. Synthetic gene circuits can have adjustable expression levels and an expression-dependent fitness peak. That means that their protein levels can be controlled to be more or less nonoptimal, adjusting how far a cell population is from the fitness peak. Evolving such cells in the lab reveals whether they move toward the fitness peak in slow steps or in jumps, as well as the mutations that mediate these fitness improvements (30).

### 6.5. Drug-Resistant Cells Can Survive Harsh-Drug Environments

Multidimensional drug fitness landscapes have been explored by using synthetic gene circuits. Increasing the concentration of an antibiotic drug (here, Zeocin) makes the environment harsher for a bacterial cell. At the same time, a small inserted gene circuit (of two genes) can be used to vary the cell's drug resistance. The level of resistance can be controlled by adding an inducer (here, anhydrotetracycline) in the medium (50). Figure 13 shows the observed fitness landscape that results from systematically controlling these two variables. It shows that harsher environments cause cells to grow slower. It also shows that turning up the drug resistance to a certain point in the cells increases the growth rate of the cell population in harsh environments. However, drug resistance is costly: Turning it on alone, by adding inducer without antibiotic, slows the cell division rate. This causes the growth rate to slow down even in antibiotic if drug resistance is turned up too high, exceeding a certain Goldilocks point beyond which it becomes too costly. This creates a crest in the landscape, given by the line connecting fitness peaks at each drug concentration. In the section below, we describe additional important—but subtle—insights that come from this experiment.

### 6.6. Protein Abundances Are Distributions, Not Single Numbers; Some Are Bimodal

A population is the collection of its cells. Take a property such as protein abundance. In some cases, the population property will have a unimodal single-peaked distribution, with a mean value that reflects that of a typical cell. In other cases, the distribution will be multimodal, with multiple peaks. In that case, the population average does not well reflect a

typical cell. For a distribution of greenish beads and reddish beads in a barrel, the average is not a yellow bead. It has recently become possible to study subpopulations of cells systematically by using synthetic gene circuits. In the following example, one variable is the environmental harshness, which, as above, is controlled by the drug Zeocin. A second variable is the drug resistance of the cell, which is controlled by the inducer concentration. Now, beyond the experiment described above, which measured the whole-population response, we also examine the distribution of drug-resistance gene expression over individual cells (50, 99). Figure 14 shows that single cells can have bimodal distributions, with peaks labeled ON or OFF. Distinct protein levels in these bimodal distributions correspond to distinct division rates of single cells,  $\lambda_c(x_{\text{ON}})$  and  $\lambda_c(x_{\text{OFF}})$ , according to their locations  $x_{\text{ON}}$  and  $x_{\text{OFF}}$  on microscopic cellular fitness landscapes  $\lambda_c(x)$ . The macroscopic cell population growth rate  $\Lambda$  is then the weighted average of single-cell division rates  $\lambda_c(x)$ , with weights taken from the protein-level distribution  $p(x)$ :

$$\Lambda = \int_0^\infty p(x)\lambda_c(x)dx \approx \lambda_c(x_{\text{ON}})p_{\text{ON}} + \lambda_c(x_{\text{OFF}})p_{\text{OFF}}. \quad 16.$$

This is how we obtain the macroscopic growth rate from microscopic single-cell division rates.

In multimodal distributions, it is not just the relative populations that matter. When gene expression is bimodal, individual cells can switch from one subpopulation to the other with rates of transition  $r_{\text{ON} \rightarrow \text{OFF}}$  and  $r_{\text{OFF} \rightarrow \text{ON}}$ . These switching rates can affect the properties of the whole population. A simple model for the switching dynamics is

$$\begin{aligned} \frac{dN_{\text{ON}}}{dt} &= [\lambda_c(x_{\text{ON}}) - r_{\text{ON} \rightarrow \text{OFF}}]N_{\text{ON}} + r_{\text{OFF} \rightarrow \text{ON}}N_{\text{OFF}}, \\ \frac{dN_{\text{OFF}}}{dt} &= [\lambda_c(x_{\text{OFF}}) - r_{\text{OFF} \rightarrow \text{ON}}]N_{\text{OFF}} + r_{\text{ON} \rightarrow \text{OFF}}N_{\text{ON}}, \end{aligned} \quad 17.$$

where the  $N$ s are the populations of the two states. The growth rate of the cell population is computed from the largest eigenvalue of this set of equations. The dwell time of a system in one state or the other is called the memory; for example,  $\tau_{\text{ON}} = \ln(2)/r_{\text{ON} \rightarrow \text{OFF}}$ . This dwell time is important because it affects the protein-level distribution, which affects the fitness in constant or fluctuating environments (5, 51). The timescale of such cellular memory in a population can be comparable to, or even longer than, the timescales of evolutionary events such as the fixation of new genotypes.

A useful insight derives from whether the cells populate a unimodal or multimodal distribution. If cells live in a constant environment, and if they have an initial bimodal distribution, they will evolve toward a unimodal distribution because that maximizes the fitness over the whole population. Mutant cells obeying this principle have indeed been observed experimentally (Figure 14) in multiple environments, indicating the generality of this principle (99).

## 7. BIG ADAPTATIONS: DYNAMICS AND MULTISTABILITY

### 7.1. When Are Static Landscapes Not Enough? The Red Queen Effect

As noted above, the general dynamics of adaptation can be expressed as two types of forces (37): potential-like forces, which are seen in systems that relax to equilibrium and are computed as downhill slopes on a potential surface, and curl-flux forces, which apply to multidimensional systems that are out of equilibrium. Wright's adaptive fitness landscapes did not account for the latter, the dynamical forces. Below are examples of where curl-flux components matter.

The Red Queen effect (100) can be viewed as a consequence of the curl flux. The term Red Queen comes from the Alice in Wonderland story. To paraphrase, sometimes you have to run just to stay in place. It describes predator-prey effects, or coevolution: Predator chases prey, prey evolves to escape the predator, and predator then evolves to better capture the prey, in a vicious cycle. In this process, neither species is necessarily changing its own individual fitness for its environment. Rather, both species can now be linked in a two-body cycle (100, 101). Figure 15 illustrates the curl flux on a landscape that has the shape of a Mexican hat. The predator-prey pair is attracted to the cycle valley, where the curl flux describes the chase around the ring (16, 22, 37). Red Queen dynamics is a multiagent multidimensional property.

The curl-flux component has profound implications for adaptation dynamics. It means that the shape alone of an adaptive landscape is not always the whole story. This indicates that evolution does not always climb straight up fitness hills or act like balls rolling downhill into valleys of fitness potential. In the conventional Wright-Fisher adaptive landscape theory, the mean fitness (averaged over all the individuals in the population) is directly related to the equilibrium population. In the old Wright-Fisher view, the largest population is of cells that have the highest mean fitness, and the selection pressure is the slope of the landscape. In contrast, this new curl-flux perspective shows that evolutionary pressure is not equal to the gradient of the landscape. The state of highest fitness is not necessarily that of the most fit predator or prey alone. Likewise, fitness bottlenecks are not those of predator or prey alone either. Curl-flux effects arise not only in single-locus and coevolving systems but also in multilocus, multiallele evolution. An example is epistasis, where one gene affects another one. Protein 1 may affect cell fitness in a particular way, and then protein 2 may compensate, leading to yet another change in protein 1. These are situations where the simpler conventional adaptive landscapes of Wright and Fisher break down (37).

Curl fluxes apply when multiple actors are coevolving. However, there are many independent-actor situations, as described throughout this review, where curl-flux dynamics is not needed. Another example where curl-flux forces are important is in the cell cycle and differentiation, as described in Sections 7.2–7.4.

### 7.2. Stem Cell Differentiation: A Bifurcating Homeostasis Landscape

During the early-stage development of an organism, a primary stem cell turns into a differentiated cell. In the 1940s, Waddington described differentiation and development through his iconic epigenetic landscape, shown in Figure 16 (19, 22, 102). Metaphorically, a

ball rolls down a single valley (the stem cell state), which bifurcates into two valleys (the differentiated state). This description is rather qualitative. Figure 16 shows the simplest quantitative realization, a two-gene circuit for differentiation. The values  $x_1$  and  $x_2$  are the concentrations of the two proteins that both self-activate and mutually repress each other's gene expression. The dynamics is given by

$$\begin{aligned}\frac{dx_1}{dt} &= \frac{ax_1^n}{S^n + x_1^n} + \frac{bS^n}{S^n + x_2^n} - kx_1 = F_1, \\ \frac{dx_2}{dt} &= \frac{ax_2^n}{S^n + x_2^n} + \frac{bS^n}{S^n + x_1^n} - kx_2 = F_2,\end{aligned}\tag{18}$$

where  $S$  represents the threshold and  $n$  represents the cooperativity on the gene regulation.

Each transcription factor has three actions: It can self-activate (at rate  $a$ ), it can repress the other (at rate  $b$ ), and it can degrade (at rate  $k$ ). In this model, the transition from a primary stem cell to its differentiated state is a change from dominance of self-activation of the transcription factor to dominance of other-repression.

Here is the updated perspective on Waddington's static differentiation model, now accounting for the curl-flux dynamics, correcting some earlier inconsistencies (19, 20, 49, 103–109). The process of human embryonic stem cell differentiation and reprogramming has been modeled in terms of 50 key biomolecules (106; see also 17, 19, 20, 22, 108). The key genes and regulatory steps for differentiation/reprogramming have been identified by global sensitivity analysis based on the landscape topography—the barriers between basins of attractions of the cell states.

Figure 16 shows two pairs of pathways: one for differentiation and the other for reprogramming (reverse differentiation). There are important differences between the quantified Waddington landscape with the dynamics-corrected model and the original picture. First, Waddington's stem cell should have been in a metastable state rather than an unstable state; its fitness potential needed to be higher than for the bifurcated states. Second, the dynamics-corrected model shows that differentiation processes are determined both by the induction driven by the programmed regulation changes along development and by the fluctuations. Differentiation is not purely spontaneous, as implied by the Waddington picture. Third, the dynamics-corrected model shows that the pathway from the multipotent to the differentiated state is different from the route of reprogramming from the differentiated state back to the stem state, in contrast to Waddington's reversible paths. This is a signature of nonequilibrium and broken-detailed-balance contributions to adaptation.

### 7.3. Understanding the Direction of Time's Arrow in Adaptive Processes Requires Accounting for Curl Fluxes

Fourth, differentiation is irreversible. The origin of the direction of time has been a puzzle in differentiation/development. This is because common descriptions of differentiation/development show no apparent signature of such direction (22, 110). Accounting for the curl flux resolves this puzzle, because it shows how detailed balance, and hence time

reversibility, is broken in these nonequilibrium states. Time's arrow points in the direction of a particular sequence of events on the curl-flux-dynamics-corrected landscape, whereas there is no direction on detailed-balanced landscapes (19, 108, 110). This dynamics-corrected picture is consistent with data from RNA sequencing experiments (111–115). In short, this modeling shows where static landscapes, and where the balls-rolling-downhill picture, are too simple to capture key dynamical aspects of the biology of differentiation and development (19).

#### 7.4. Cell Cycles and Regulation Are Also Subject to Curl Fluxes

Curl-flux dynamics is also needed to account for cell cycle dynamics and cell regulation (16, 116). The cell cycle has states  $G_1$ , S,  $G_2$ , and M, which are landscape basins on the cycle pathway. Cycle checkpoints are the transition states (locations of the barriers) between these local basins that affect go/no go decisions (16, 116). The cell cycle pathway combines (a) a Mexican hat landscape, where the system is stable anywhere along the hat brim, with (b) the curl-flux dynamics of cell cycle oscillations (running around the brim). This has also been described in the context of evolution with the group-help model (37). As a matter of principle in evolution, the adaptation rate—when curl-flux dynamics is taken into account—is found to be dependent not purely on the genetic variance but also on interactions between evolving agents. This point is a generalization of Fisher's fundamental theorem of natural selection (37), which states that the rate of mean fitness increase of an organism at a specific time is proportional to its genetic variance at that time (117). In short, even when evolution reaches fitness maxima, where the fitness cannot increase more, evolution can still continue: Agents can evolve further since the genetic variance is still driven by an intrinsic flux such as coevolution of predator and prey (37), in contrast with the traditional view that only fitness differences can drive evolution.

## 8. SUMMARY

Cells experience adaptive forces. Cells undergo homeostasis. Populations undergo evolution. These adaptive actions can be expressed in terms of driving forces and landscapes. Here, we review modeling that is quantitative, that is tested or testable, and that addresses more physical and combinatoric properties of the genotype-to-phenotype mapping problem than sequence comparisons alone are likely to reveal. For example, adaptation depends on protein expression levels, on folding and aggregation, and on physical transport and mechanical properties. Such properties are often not explainable from pinpoints in the genome; they are delocalized and combinatoric. We review how these forces contribute to cellular growth laws. Such models provide a way of knowing that is based on hypothesizing DOFs and fitness functions, making predictions, and testing experimentally.

## ACKNOWLEDGMENTS

We appreciate the support of the Laufer Center for Physical and Quantitative Biology, the National Institutes of Health (NIH), and the National Science Foundation (NSF). In particular, L.A. is supported by the NIH (R01 GM125813), G.B. is supported by the NIH/National Institute of General Medical Sciences Maximizing Investigators' Research Award Program (R35 GM122561), and J.W. is supported by the NSF (PHY-76066 and CHE-1808474) and the NIH (R01 GM124177). We are grateful to Sasha Levy for some early influential discussions.

## LITERATURE CITED

1. Blanco C, Janzen E, Pressman A, Saha R, Chen IA. 2019 Molecular fitness landscapes from high-coverage sequence profiling. *Annu. Rev. Biophys* 48:1–18 [PubMed: 30601678]
2. Kussell E 2013 Evolution in microbes. *Annu. Rev. Biophys* 42:493–514 [PubMed: 23654305]
3. de Vos MG, Poelwijk FJ, Tans SJ. 2013 Optimality in evolution: new insights from synthetic biology. *Curr. Opin. Biotechnol* 24:797–802 [PubMed: 23684729]
4. Poelwijk FJ, De Vos MG, Tans SJ. 2011 Tradeoffs and optimality in the evolution of gene regulation. *Cell* 146:462–70 [PubMed: 21802129]
5. Acar M, Mettetal JT, van Oudenaarden A. 2008 Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet* 40:471–75 [PubMed: 18362885]
6. Barrick JE, Lenski RE. 2013 Genome dynamics during experimental evolution. *Nat. Rev. Genet* 14:827–39 [PubMed: 24166031]
7. Levy SF, Blundell JR, Venkataram S, Petrov DA, Fisher DS, Sherlock G. 2015 Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* 519:181–86 [PubMed: 25731169]
8. Van Kampen NG. 1992 *Stochastic Processes in Physics and Chemistry*, Vol. 1 Amsterdam:Elsevier 2nd ed.
9. Hu G 1994 *Stochastic Forces and Nonlinear Systems*. Shanghai: Shanghai Sci. Technol. Educ.
10. Gardiner CW. 1985 *Handbook of Stochastic Methods*, Vol. 3 Berlin: Springer
11. Risken H 1996 Fokker-Planck equation In *The Fokker-Planck Equation*, pp. 63–95. Berlin: Springer
12. Graham R 1989 Macroscopic potentials, bifurcations and noise in dissipative systems In *Noise in Nonlinear Dynamical Systems*, Vol. 1: Theory of Continuous Fokker–Planck Systems, ed. Moss F, McClintock PVE, pp. 225–78. Cambridge, UK: Cambridge Univ. Press
13. Sasai M, Wolynes PG. 2003 Stochastic gene expression as a many-body problem. *PNAS* 100:2374–79 [PubMed: 12606710]
14. Qian H 2009 Entropy demystified: the “thermo”-dynamics of stochastically fluctuating systems. *Methods Enzymol.* 467:111–34 [PubMed: 19897091]
15. Ao P 2008 Emerging of stochastic dynamical equalities and steady state thermodynamics from Darwinian dynamics. *Commun. Theor. Phys* 49:1073–90 [PubMed: 21949462]
16. Wang J, Xu L, Wang E. 2008 Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *PNAS* 105:12271–76 [PubMed: 18719111]
17. Wang J, Zhang K, Wang E. 2010 Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *J. Chem. Phys* 133:09B613
18. Feng H, Wang J. 2011 Potential and flux decomposition for dynamical systems and nonequilibrium thermodynamics: curvature, gauge field, and generalized fluctuation-dissipation theorem. *J. Chem. Phys* 135:234511 [PubMed: 22191890]
19. Wang J, Zhang K, Xu L, Wang E. 2011 Quantifying the Waddington landscape and biological paths for development and differentiation. *PNAS* 108:8257–62 [PubMed: 21536909]
20. Feng H, Zhang K, Wang J. 2014 Non-equilibrium transition state rate theory. *Chem. Sci* 5:3761–69
21. Qian H 2014 The zeroth law of thermodynamics and volume-preserving conservative system in equilibrium with stochastic damping. *Phys. Lett. A* 378:609–16
22. Wang J 2015 Landscape and flux theory of non-equilibrium dynamical systems with application to biology. *Adv. Phys* 64:1–137
23. Shav-Tal Y, Singer RH, Darzacq X. 2004 Imaging gene expression in single living cells. *Nat. Rev. Mol. Cell Biol* 5:855–61 [PubMed: 15459666]
24. Gardner TS, Cantor CR, Collins JJ. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–42 [PubMed: 10659857]
25. Elowitz MB, Leibler S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–38 [PubMed: 10659856]

26. Becskei A, Serrano L. 2000 Engineering stability in gene networks by autoregulation. *Nature* 405:590–93 [PubMed: 10850721]
27. Levskaya A, Chevalier AA, Tabor JJ, Simpson ZB, Lavery LL, et al. 2005 Engineering *Escherichia coli* to see light. *Nature* 438:441–42 [PubMed: 16306980]
28. Hooshangi S, Thiberge S, Weiss R. 2005 Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *PNAS* 102:3581–86 [PubMed: 15738412]
29. Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J. 2008 A fast, robust and tunable synthetic gene oscillator. *Nature* 456:516–19 [PubMed: 18971928]
30. Gonzalez C, Ray JC, Manhart M, Adams RM, Nevozhay D, et al. 2015 Stress-response balance drives the evolution of a network module and its host genome. *Mol. Syst. Biol* 11:827 [PubMed: 26324468]
31. Poelwijk FJ, de Vos MG, Tans SJ. 2011 Tradeoffs and optimality in the evolution of gene regulation. *Cell* 146:462–70 [PubMed: 21802129]
32. Farquhar KS, Charlebois DA, Szenk M, Cohen J, Nevozhay D, Balazsi G. 2019 Role of network-mediated stochasticity in mammalian drug resistance. *Nat. Commun* 10:2766 [PubMed: 31235692]
33. Sleight SC, Bartley BA, Lieviant JA, Sauro HM. 2010 Designing and engineering evolutionary robust genetic circuits. *J. Biol. Eng* 4:12 [PubMed: 21040586]
34. Ferea TL, Botstein D, Brown PO, Rosenzweig RF. 1999 Systematic changes in gene expression patterns following adaptive evolution in yeast. *PNAS* 96:9721–26 [PubMed: 10449761]
35. Lind PA, Libby E Herzog J, Rainey PB. 2019 Predicting mutational routes to new adaptive phenotypes. *eLife* 8:e38822 [PubMed: 30616716]
36. Arnold FH. 2015 The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys* 48:404–10 [PubMed: 26537398]
37. Zhang F, Xu L, Zhang K, Wang E, Wang J. 2012 The potential and flux landscape theory of evolution. *J. Chem. Phys* 137:065102 [PubMed: 22897313]
38. Zaňudo JGT, Guinn MT, Farquhar K, Szenk M, Steinway SN, et al. 2004 Towards control of cellular decision-making networks in the epithelial-to-mesenchymal transition. *Phys. Biol* 16:031002
39. MacNeil LT, Walhout AJM. 2011 Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* 21:645–57 [PubMed: 21324878]
40. Wright S 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution In *Proceedings of the Sixth International Congress of Genetics*, Vol. 1, ed. Jones DF, pp. 356–66. Menasha, WI: Brooklyn Bot. Gard.
41. Kuchner O, Arnold FH. 1997 Directed evolution of enzyme catalysts. *Trends Biotechnol.* 15:523–30 [PubMed: 9418307]
42. Jiménez JI, Xulvi-Brunet R, Campbell GW, Turk-MacLeod R, Chen IA. 2013 Comprehensive experimental fitness landscape and evolutionary network for small RNA. *PNAS* 110:14984–89 [PubMed: 23980164]
43. Sailer ZR, Harms MJ. 2017 Molecular ensembles make evolution unpredictable. *PNAS* 114:11938–43 [PubMed: 29078365]
44. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. 2007 Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* 445:383–86 [PubMed: 17251971]
45. Harms MJ, Thornton JW. 2014 Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512:203–7 [PubMed: 24930765]
46. Sailer ZR, Harms MJ. 2017 High-order epistasis shapes evolutionary trajectories. *PLOS Comput. Biol* 13:e1005541 [PubMed: 28505183]
47. Aharoni A, Gaidukov L, Khersonsky O, Gould SM, Roodveldt C, Tawfik DS. 2005 The “evolability” of promiscuous protein functions. *Nat. Genet* 37:73–76 [PubMed: 15568024]
48. Becskei A, Séraphin B, Serrano L. 2001 Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* 20:2528–35 [PubMed: 11350942]

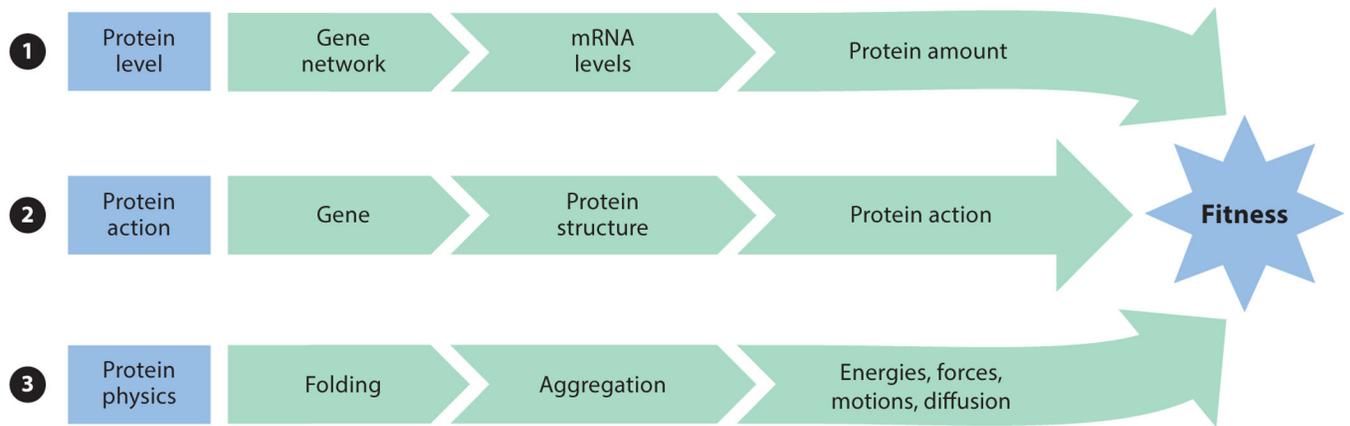
49. Ferrell JE Jr. 2012 Bistability, bifurcations, and Waddington's epigenetic landscape. *Curr. Biol* 22:R458–66 [PubMed: 22677291]
50. Nevozhay D, Adams RM, Van Itallie E, Bennett MR, Balázsi G. 2012 Mapping the environmental fitness landscape of a synthetic gene circuit. *PLOS Comput. Biol* 8:e1002480 [PubMed: 22511863]
51. Acar M, Becskei A, van Oudenaarden A. 2005 Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435:228–32 [PubMed: 15889097]
52. Isaacs FJ, Hasty J, Cantor CR, Collins JJ. 2003 Prediction and measurement of an autoregulatory genetic module. *PNAS* 100:7714–19 [PubMed: 12808135]
53. Jackson EA. 1992 *Perspectives of Nonlinear Dynamics, Vol. 1* Cambridge, UK: Cambridge Univ. Press Arch.
54. Wang J, Li C, Wang E. 2010 Potential and flux landscapes quantify the stability and robustness of budding yeast cell cycle network. *PNAS* 107:8195–200 [PubMed: 20393126]
55. Wu W, Wang J. 2014 Potential and flux field landscape theory. II. Non-equilibrium thermodynamics of spatially inhomogeneous stochastic dynamical systems. *J. Chem. Phys* 141:105104 [PubMed: 25217956]
56. Xu L, Shi H, Feng H, Wang J. 2012 The energy pump and the origin of the non-equilibrium flux of the dynamical systems and the networks. *J. Chem. Phys* 136:165102 [PubMed: 22559506]
57. Xu L, Zhang F, Wang E, Wang J. 2013 The potential and flux landscape, Lyapunov function and non-equilibrium thermodynamics for dynamic systems and networks with an application to signal-induced  $\text{Ca}^{2+}$  oscillation. *Nonlinearity* 26:R69
58. Xu L, Zhang F, Zhang K, Wang E, Wang J. 2014 The potential and flux landscape theory of ecology. *PLOS ONE* 9:e86746 [PubMed: 24497975]
59. Sella G, Hirsh AE. 2005 The application of statistical physics to evolutionary biology. *PNAS* 102:9541–46 [PubMed: 15980155]
60. Barton N, Coe J. 2009 On the application of statistical physics to evolutionary biology. *J. Theor. Biol* 259:317–24 [PubMed: 19348811]
61. de Vladar HP, Barton NH. 2011 The contribution of statistical physics to evolutionary biology. *Trends Ecol. Evol* 26:424–32 [PubMed: 21571390]
62. Echave J, Wilke CO. 2017 Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. *Annu. Rev. Biophys* 46:85–103 [PubMed: 28301766]
63. Agozzino L, Dill KA. 2018 Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *PNAS* 115:9092–97 [PubMed: 30150386]
64. Orr HA. 2009 Fitness and its role in evolutionary genetics. *Nat. Rev. Genet* 10:531–39 [PubMed: 19546856]
65. Orr HA. 2015 The genetic theory of adaptation: a brief history. *Nat. Rev. Genet* 6:119–27
66. Sawle L, Ghosh K. 2011 How do thermophilic proteins and proteomes withstand high temperature? *Biophys. J* 101:217–27 [PubMed: 21723832]
67. Dill KA, Ghosh K, Schmit JD. 2011 Physical limits of cells and proteomes. *PNAS* 108:17876–82 [PubMed: 22006304]
68. Ghosh K, de Graff AM, Sawle L, Dill KA. 2016 Role of proteome physical chemistry in cell behavior. *J. Phys. Chem. B* 120:9549–63 [PubMed: 27513457]
69. Levy SF, Ziv N, Siegal ML. 2012 Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLOS Biol.* 1:e1001325
70. Charlebois DA, Hauser K, Marshall S, Balazsi G. 2018 Multiscale effects of heating and cooling on genes and gene networks. *PNAS* 115:E10797–806 [PubMed: 30341217]
71. Blaby IK, Lyons BJ, Wroclawska-Hughes E, Phillips GC, Pyle TP, et al. 2012 Experimental evolution of a facultative thermophile from a mesophilic ancestor. *Appl. Environ. Microbiol* 78:144–55 [PubMed: 22020511]
72. Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005 Why highly expressed proteins evolve slowly. *PNAS* 102:14338–43 [PubMed: 16176987]
73. Zeldovich KB, Chen P, Shakhnovich EI. 2007 Protein stability imposes limits on organism complexity and speed of molecular evolution. *PNAS* 104:16152–57 [PubMed: 17913881]

74. Drummond DA, Wilke CO. 2008 Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* 134:341–52 [PubMed: 18662548]
75. Yang JR, Zhuang SM, Zhang J. 2010 Impact of translational error-induced and error-free misfolding on the rate of protein evolution. *Mol. Syst. Biol* 6:421 [PubMed: 20959819]
76. Serohijos AWR, Lee SYR, Shakhnovich EI. 2013 Highly abundant proteins favor more stable 3D structures in yeast. *Biophys. J* 104:L1–3 [PubMed: 23442924]
77. Sikosek T, Chan HS. 2014 Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* 11:20140419 [PubMed: 25165599]
78. Serohijos AWR, Rimas Z, Shakhnovich EI. 2012 Protein biophysics explains why highly abundant proteins evolve slowly. *Cell Rep.* 2:249–56 [PubMed: 22938865]
79. Yang JR, Liao BY, Zhuang SM, Zhang J. 2012 Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *PNAS* 109:E831–40 [PubMed: 22416125]
80. Levy ED, De S, Teichmann SA. 2012 Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *PNAS* 109:20461–66 [PubMed: 23184996]
81. Wylie CS, Shakhnovich EI. 2011 A biophysical protein folding model accounts for most mutational fitness effects in viruses. *PNAS* 108:9916–21 [PubMed: 21610162]
82. Santra M, Farrell DW, Dill KA. 2017 Bacterial proteostasis balances energy and chaperone utilization efficiently. *PNAS* 114:E2654–61 [PubMed: 28292901]
83. Maitra A, Dill KA. 2015 Bacterial growth laws reflect the evolutionary importance of energy efficiency. *PNAS* 112:406–11 [PubMed: 25548180]
84. Maitra A, Dill KA. 2016 Modeling the overproduction of ribosomes when antibacterial drugs act on cells. *Biophys. J* 110:743–48 [PubMed: 26840738]
85. Wagoner JA, Dill KA. 2019 Mechanisms for achieving high speed and efficiency in biomolecular machines. *PNAS* 116:5902–7 [PubMed: 30850521]
86. Wagoner JA, Dill KA. 2019 Opposing pressures of speed and efficiency guide the evolution of molecular machines. *Mol. Biol. Evol* 36:2813–22 [PubMed: 31432071]
87. Schuech R, Hoehfurtner T, Smith DJ, Humphries S. 2019 Motile curved bacteria are Pareto-optimal. *PNAS* 116:14440–47 [PubMed: 31266896]
88. Eames M, Kortemme T. 2012 Cost-benefit tradeoffs in engineered *lac* operons. *Science* 336:911–15 [PubMed: 22605776]
89. Perfeito L, Ghozzi S, Berg J, Schnetz K, Lässig M. 2011 Nonlinear fitness landscape of a molecular pathway. *PLOS Genet.* 7:e1002160 [PubMed: 21814515]
90. Nevozhay D, Adams RM, Van Itallie E, Bennett MR, Balazsi G. 2012 Mapping the environmental fitness landscape of a synthetic gene circuit. *PLOS Comput. Biol* 8:e1002480 [PubMed: 22511863]
91. Tan C, Marguet P, You L. 2009 Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol* 5:842–48 [PubMed: 19801994]
92. Gill G, Ptashne M. 1988 Negative effect of the transcriptional activator GAL4. *Nature* 334:721–24 [PubMed: 3412449]
93. Nikolados EM, Weiße AY, Ceroni F, Oyarzun DA. 2019 Growth defects and loss-of-function in synthetic gene circuits. *ACS Synth. Biol* 8:1231–40 [PubMed: 31181895]
94. Deris JB, Kim M, Zhang Z, Okano H, Hermsen R, et al. 2013 The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* 342:1237435 [PubMed: 24288338]
95. Hasty J, McMillen D, Collins JJ. 2002 Engineered gene circuits. *Nature* 420:224–30 [PubMed: 12432407]
96. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009 Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol* 5:593–99 [PubMed: 19561621]
97. Goldsmith M, Tawfik DS. 2017 Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol* 47:140–50 [PubMed: 29035814]
98. Ray JC, Wickersheim ML, Jalihal AP, Adeshina YO, Cooper TF, Balazsi G. 2016 Cellular growth arrest and persistence from enzyme saturation. *PLOS Comput. Biol* 12:e1004825 [PubMed: 27010473]

99. González C, Ray JCJ, Manhart M, Adams RM, Nevozhay D, et al. 2015 Stress-response balance drives the evolution of a network module and its host genome. *Mol. Syst. Biol* 11:827 [PubMed: 26324468]
100. Van Valen L 1973 A new evolutionary law. *Evol. Theory* 1:1–30
101. Rice SH. 2004 *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sunderland, MA: Sinauer
102. Waddington CH. 2014 (1957). *The Strategy of the Genes*. London: Routledge
103. Zhang B, Wolynes PG. 2015 Topology, structures, and energy landscapes of human chromosomes. *PNAS* 112:6062–67 [PubMed: 25918364]
104. Huang S, Guo Y, May G, Enver T. 2007 Bifurcation dynamics of cell fate decision in bipotent progenitor cells. *Dev. Biol* 305:695–713 [PubMed: 17412320]
105. Feng H, Wang J. 2012 A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Sci. Rep* 2:550 [PubMed: 22870379]
106. Li C, Wang J. 2013 Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: landscape and biological paths. *PLOS Comput. Biol* 9:e1003165 [PubMed: 23935477]
107. Li C, Wang J. 2013 Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface* 10:20130787 [PubMed: 24132204]
108. Xu L, Zhang K, Wang J. 2014 Exploring the mechanisms of differentiation, dedifferentiation, reprogramming and transdifferentiation. *PLOS ONE* 9:e105216 [PubMed: 25133589]
109. Huang S 2009 Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* 31:546–60 [PubMed: 19319911]
110. Wang J, Xu L, Wang E, Huang S. 2010 The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. *Biophys. J* 99:29–39 [PubMed: 20655830]
111. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. 2018 Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* 360:981–87 [PubMed: 29700229]
112. Briggs JA, Weinreb C, Wagner DE, Megason S, Peshkin L, et al. 2018 The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* 360:eaar5780 [PubMed: 29700227]
113. Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018 Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360:eaar3131 [PubMed: 29700225]
114. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, et al. 2019 Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* 176:928–43 [PubMed: 30712874]
115. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, et al. 2018 RNA velocity of single cells. *Nature* 560:494–98 [PubMed: 30089906]
116. Li C, Wang J. 2014 Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *PNAS* 111:14130–35 [PubMed: 25228772]
117. Fisher R 1958 (1930). *The Genetical Theory of Natural Selection*. Oxford, UK: Clarendon

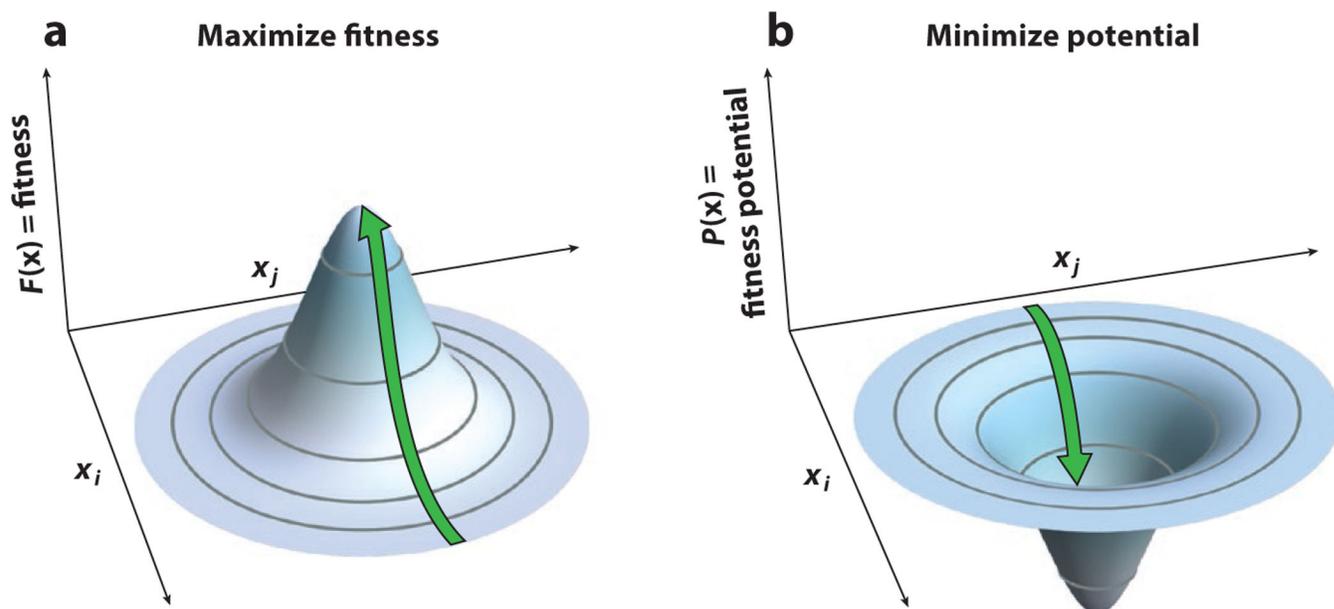
**SUMMARY POINTS**

1. New models are emerging for cell adaptation—homeostasis and evolution.
2. Insights come from postulating how fitness depends on degrees of freedom, visualizing them on landscapes, and testing them in experiments.
3. Adaptation has two kinds of forces: slopes of a potential function and curl fluxes. Curl fluxes explain important dynamical multibody effects, such as predator-prey relations.
4. Learning the cell adaptation code requires knowing mechanisms, not just comparing sequences, because fitness depends on high-order combinatorics, for example, in the folding stabilities, aggregation, abundance levels, or diffusional transport of proteins in the cell.

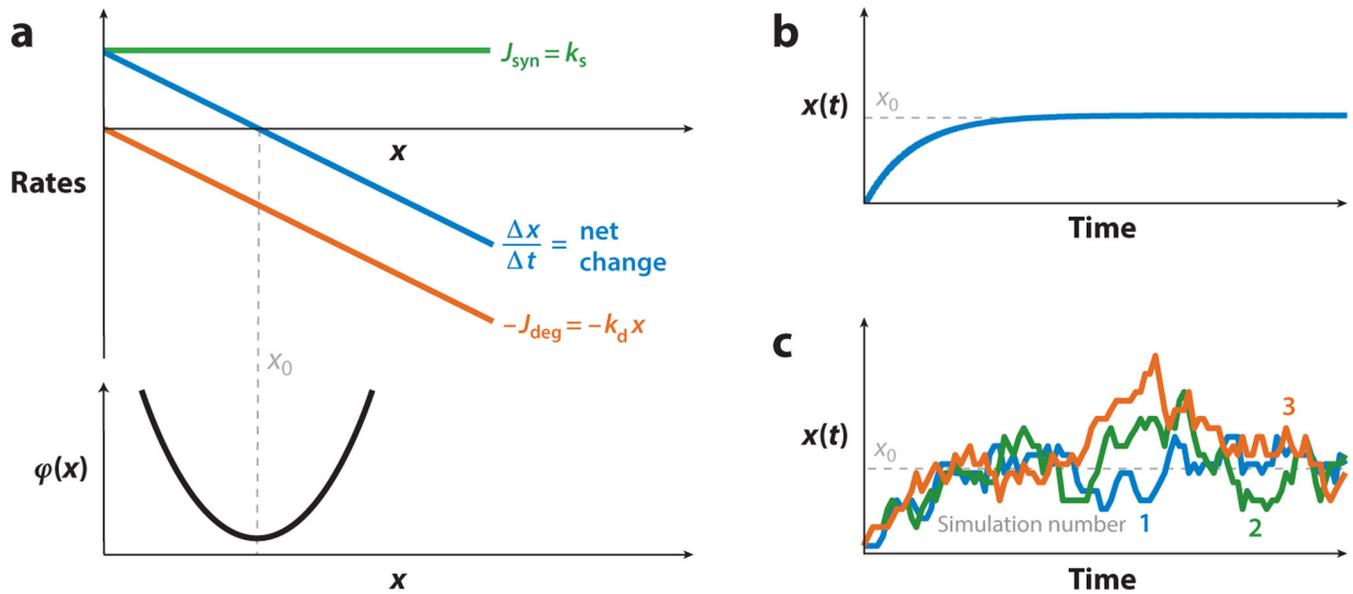


**Figure 1.**

Three ways a cell's fitness is encoded in its proteins: (❶) its abundance in the cell, affected by messenger RNA (mRNA) levels; (❷) its efficacy of biological action, affected by mutations; and (❸) proteostasis, its folding and aggregation health, controlled by protein synthesis, degradation, and chaperoning.



**Figure 2.** Evolution is described as populations moving on landscapes. This is represented in two different ways, as a tendency toward either (a) maxima on a landscape of fitness or (b) minima on a landscape of fitness potential. They are just different ways to visualize the same process.

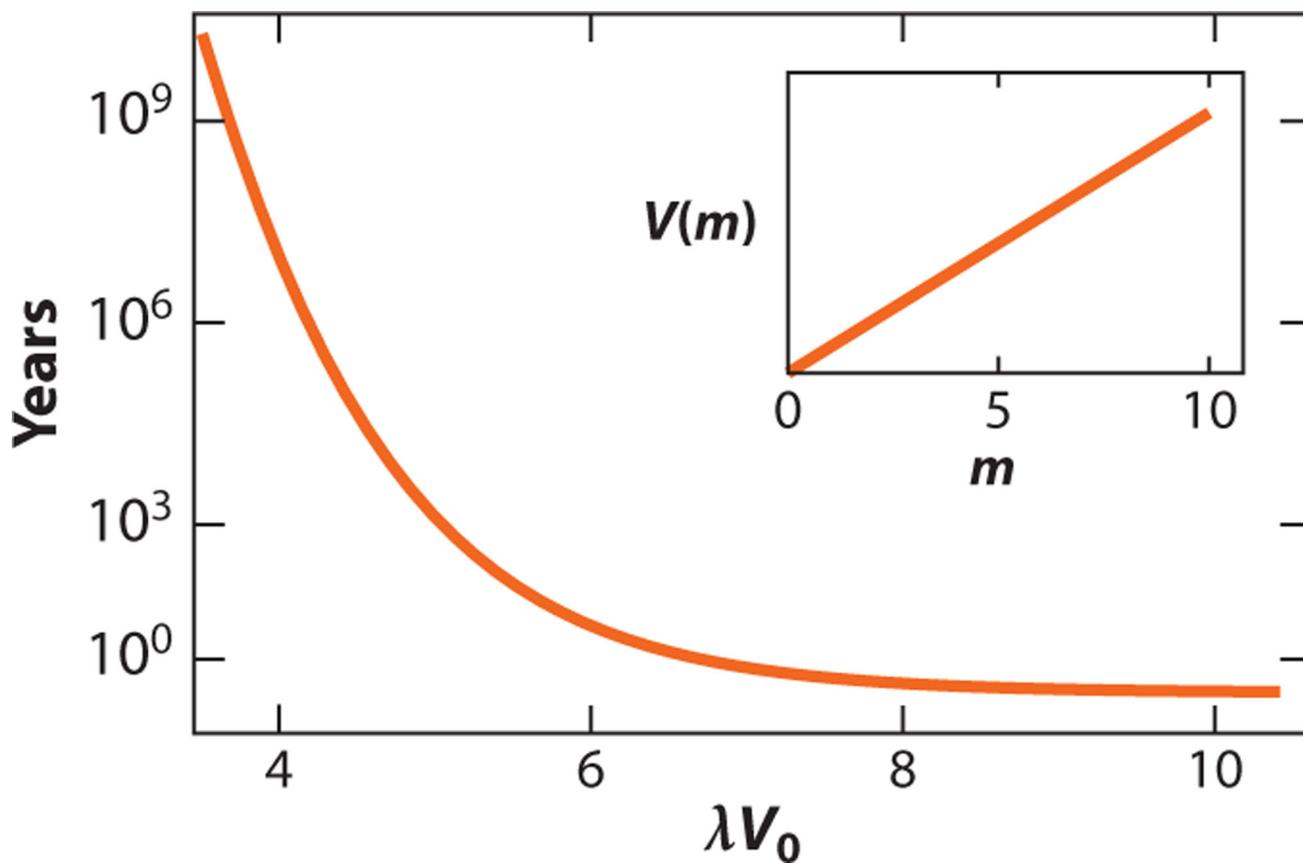


**Figure 3.** Homeostasis is a tendency toward the minimum of a potential function. (*a, top*) Homeostasis is maintained by a balance of two rates: synthesis ( $J_{\text{syn}}$ ), supplying material and increasing the concentration  $x$ , and degradation ( $J_{\text{deg}}$ ), decreasing  $x$ . The sum is the net rate of change,  $x'/t$ . (*a, bottom*) The integrated net change rate is a potential  $\phi(x)$ . (*b*) After perturbation,  $x(t)$  relaxes to  $x_0$  over time. (*c*) The noisy stochastic version of this relaxation is shown, for example, for few-particle systems.

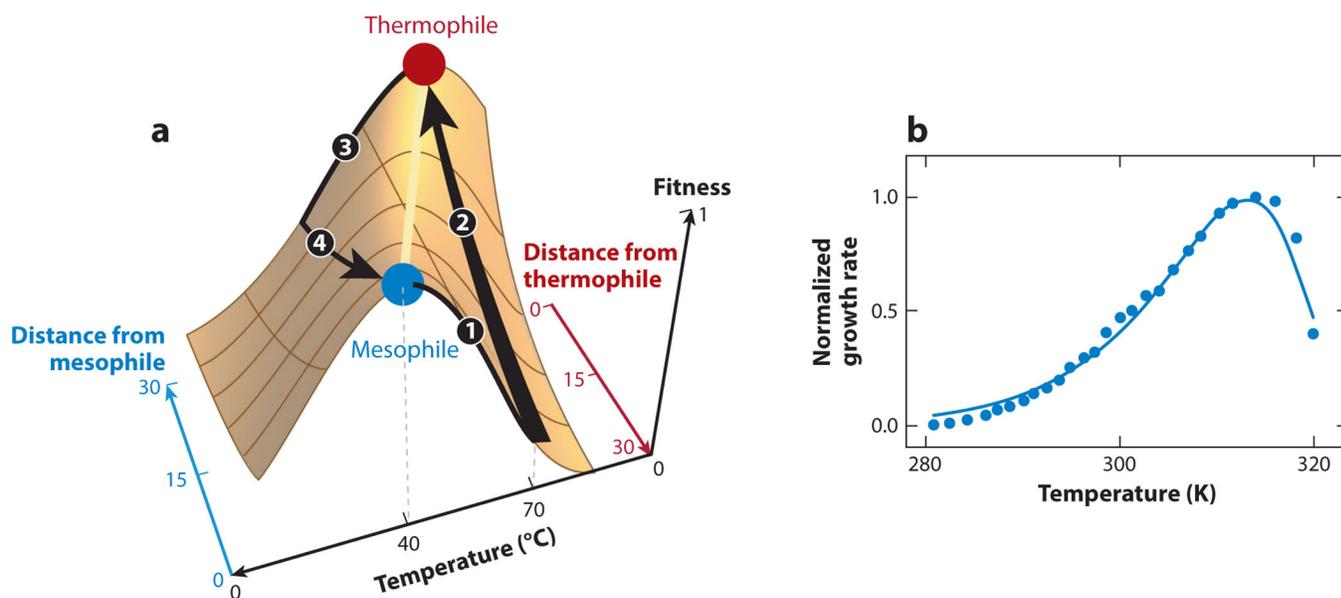
	$x$	$F(x)$	$\xi(x, t)$	Examples
<b>Cellular homeostasis</b>	Protein level	Protein interaction	Intracellular noise	$\frac{dx}{dt} = \frac{ax}{S+x} + \frac{bS}{S+y} - kx + \xi(x, y, t)$
<b>Evolutionary dynamics</b>	Allele frequency	Mutation and selection	Genetic drift	$\frac{dx}{dt} = \frac{w(x) - w_0}{w_0} x + m(y-x) + \xi(x, y, t)$

**Figure 4.**

Homeostasis dynamics and evolutionary dynamics often use similar math, but with different variables. (*Top row*) Gene module with two self-activating genes ( $x$  and  $y$ ), mutually repressing each other ( $a$  and  $b$  are the strengths of self-activation and mutual repression, respectively;  $S$  is the minimal concentration needed to activate changes; and  $k$  is the degradation rate). (*Bottom row*) Allele frequency changes due to natural selection and random mutation (22) ( $w$  is the fitness of allele  $x$ ,  $w_0$  is the average fitness, and  $m$  is the mutation rate between alleles  $x$  and  $y$ ).

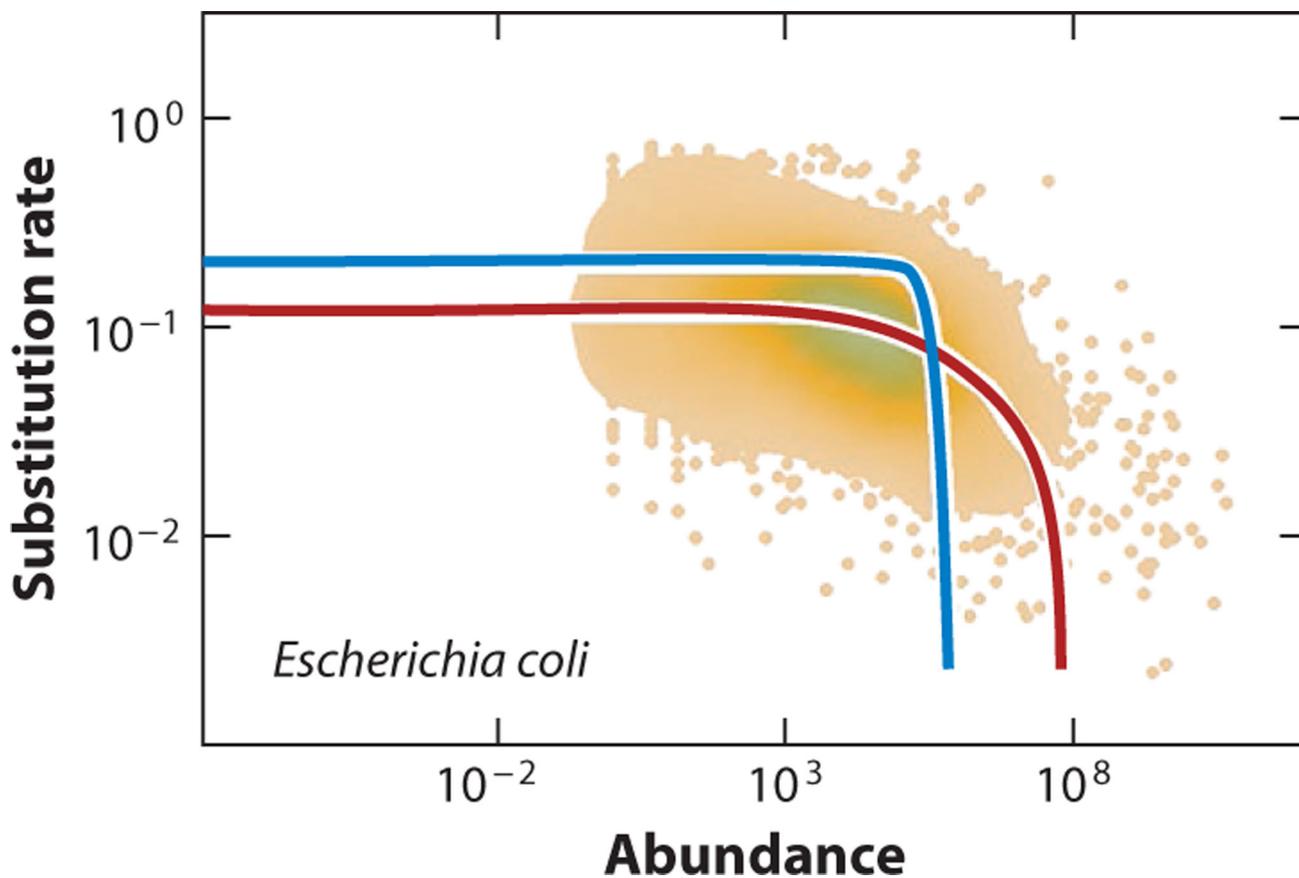


**Figure 5.** Evolution happens over a large dynamic range of timescales, modeled with the simplest (linear) fitness potential (*inset*),  $V(m) = \text{constant} \times m$  (63).



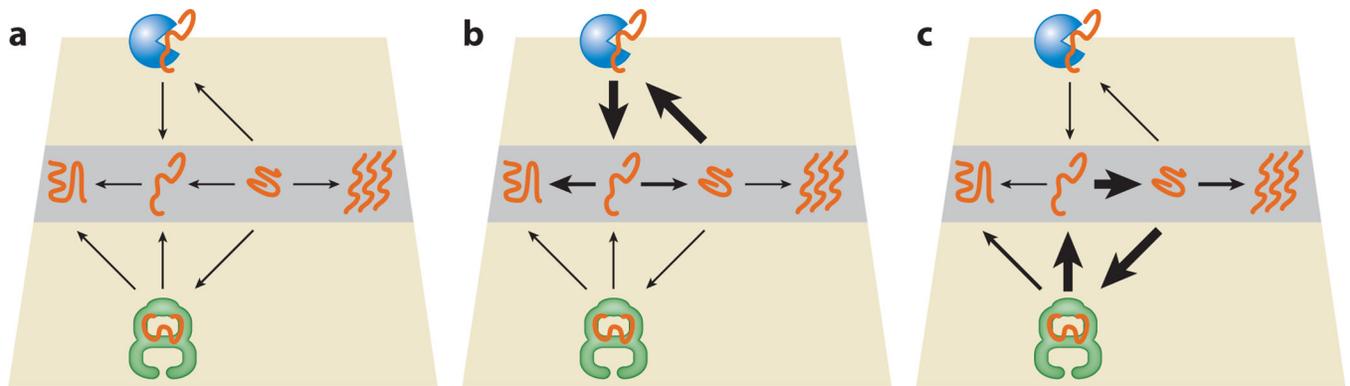
**Figure 6.**

(a) Fitness-landscape pathways for how cells evolve under changing temperatures. Route ①–② shows a (mesophilic) cell evolving to adapt to a warmer climate. Route ③–④ shows a (thermophilic) cell evolving to adapt to a colder climate. The thickness of the black arrows shows the adaptation speed computed from the thermal proteome unfolding model, which predicts that cells can adapt much faster to warmer climates than to colder ones (63). (b) Bacterial growth rates versus temperature. Panel *b* adapted with permission from Reference 68.



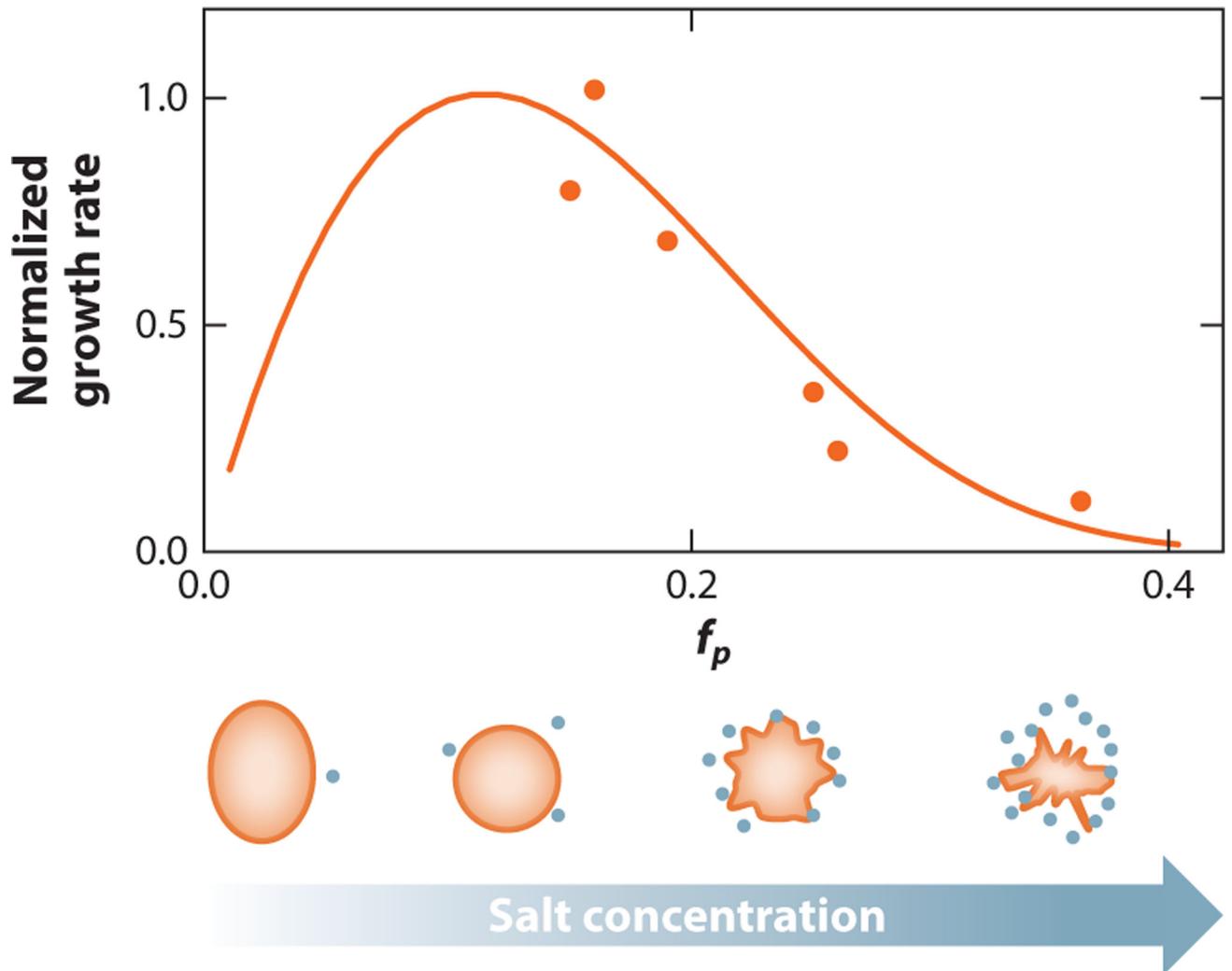
**Figure 7.**

Proteins that are most abundant in the cell are slower to evolve; that is, they have a lower rate of amino acid substitutions. A cell's fitness is more affected by mutating an abundant protein than by mutating a less abundant protein. There is a larger fitness cost to the cell for misfolding and aggregation, so the number of viable mutant sequences is smaller. The model predicts the roles of misfolding (*red line*) and of aggregation (*blue line*) for this anticorrelation between the evolution rate (given by the percentage sequence difference between orthologous proteins of related species) and abundance (measured by relative microRNA concentration) (63).

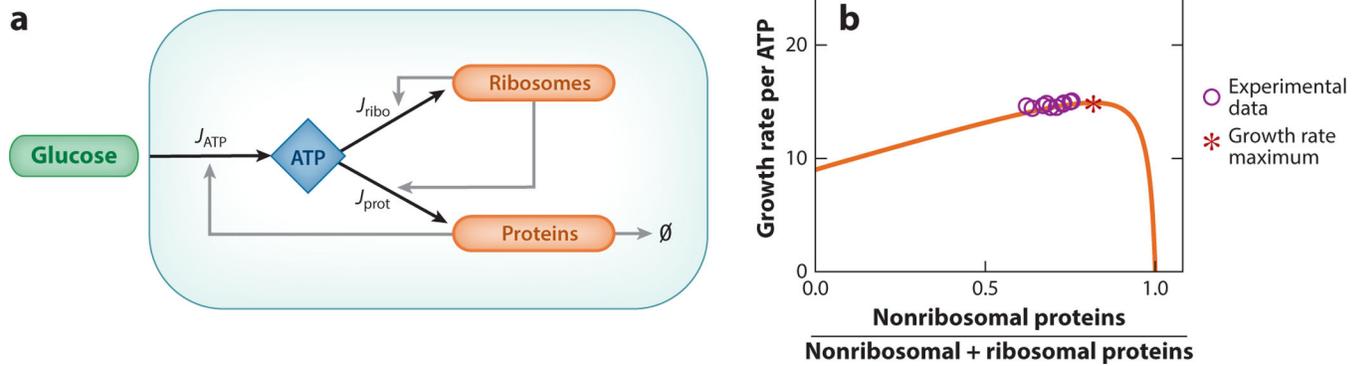


**Figure 8.**

The hospital model of proteostasis in bacteria. (a) The gray lane represents the pathway of protein folding, misfolding, and aggregation without chaperones. The other arrows show the proteostasis trafficking through different chaperones. (b) Hospital model predictions of proteostasis flows for a class II protein (mildly misfolded) indicate that it traffics mainly through the DnaK system. Heavy arrows show the main flux. (c) Hospital model predictions of proteostasis flows for a class III protein (strongly misfolded) indicate that it traffics mainly through the GroEl system.

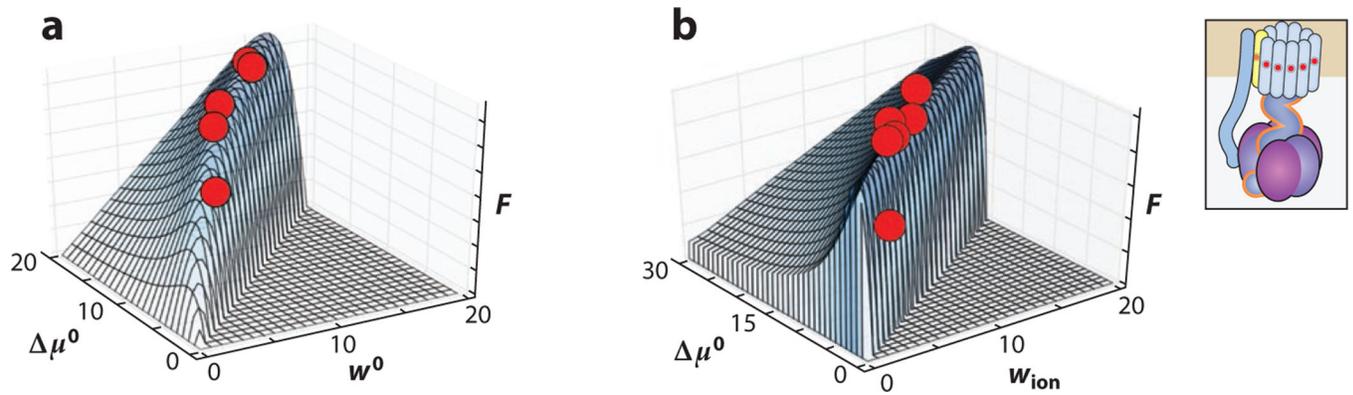


**Figure 9.** Adding external salt shrivels a cell osmotically, which increases internal protein crowding and slows protein diffusion, thus slowing cell growth. In the protein transport rate model, added salt in the surroundings ( $f_p$ ; *horizontal axis*) reduces the cell growth rate (*vertical axis*) by densifying the proteins inside and slowing their diffusional transport (68).



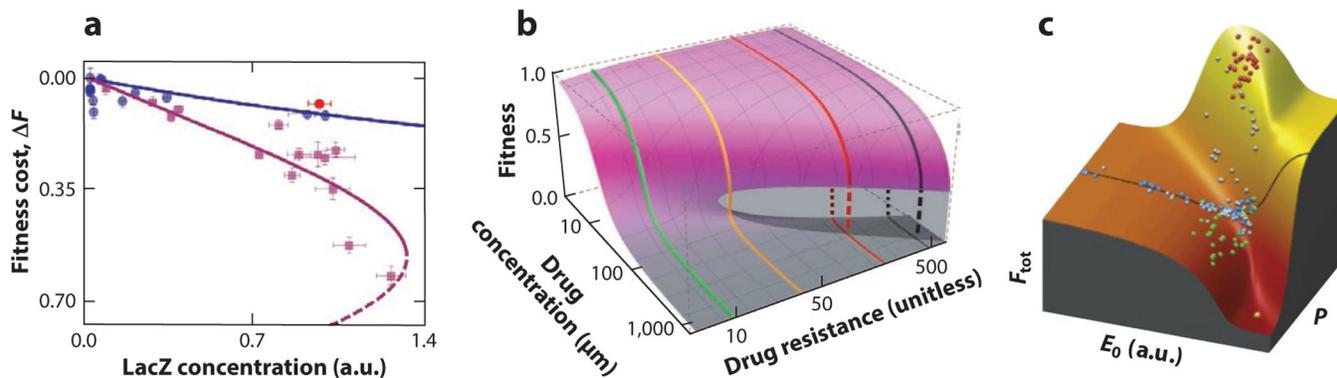
**Figure 10.**

Bacteria trade off producing ribosomes versus nonribosomal proteins. This trade-off maximizes energy efficiency. (a) In the ribosomal upswitch model,  $J_{ATP}$  is the rate of converting glucose to the nucleoside triphosphates,  $J_{ribo}$  is the production rate of ribosomal proteins, and  $J_{prot}$  is the production rate of nonribosomal proteins. These relative flows are determined by the abundance of glucose. (b) In the predicted fitness landscape, the fast-growth energy efficiency is maximized when the fraction of nonribosomal proteins is about 75% (83).



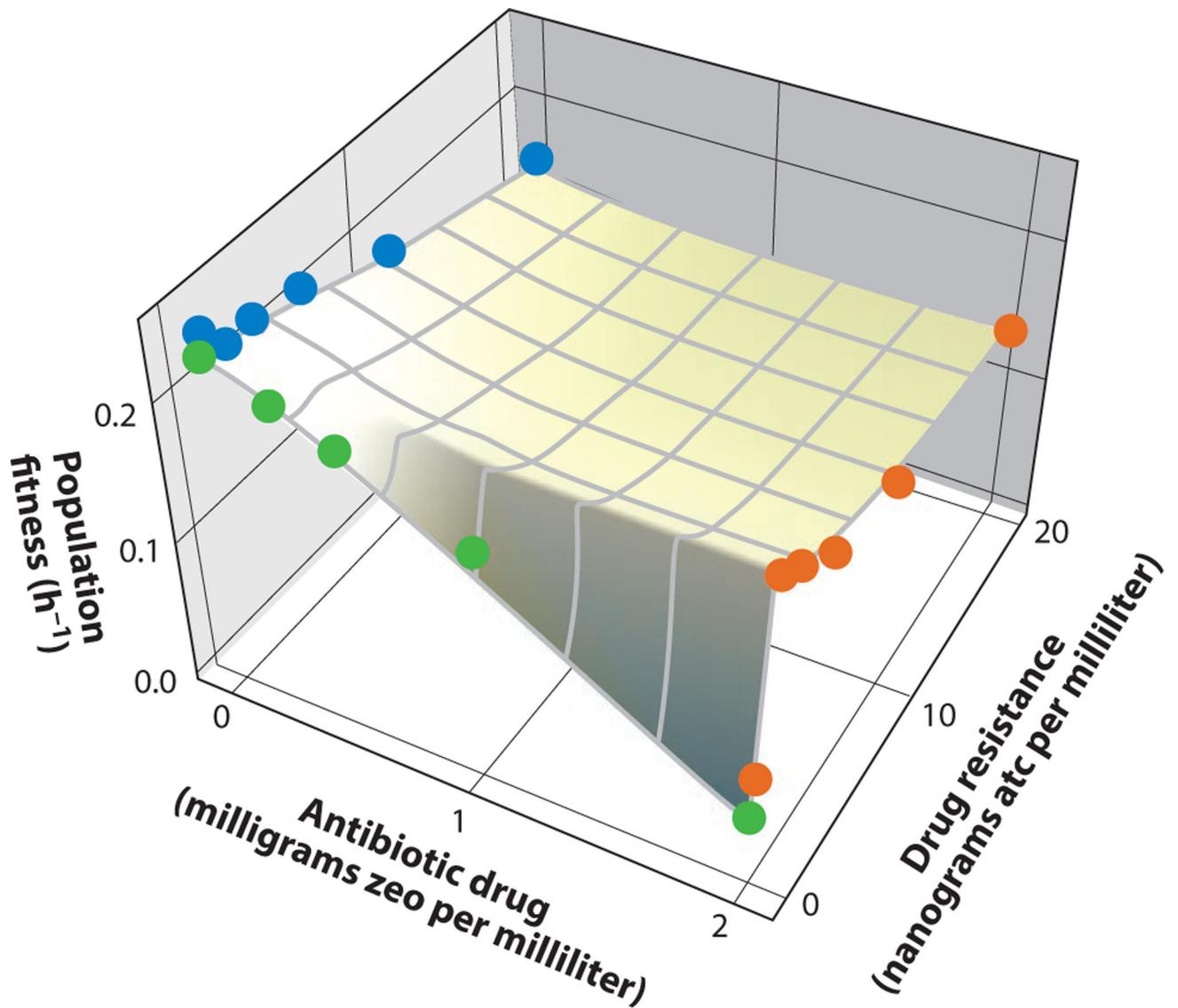
**Figure 11.**

Fitness landscapes for (a) molecular motors and (b) ion pumps. A simple model asserts a fitness function for biomolecular machines of power output per unit energy input. Panel a shows that five different motors (*red circles*) appear to optimize their output work (for a given input chemical potential from ATP degradation) by how the rate barriers are distributed through the kinetic cycle. Panel b shows the same for six different ion pumps (*red circles*). The inset shows a diagram of the  $F_0F_1$  ATPase motor. Figure adapted with permission from Reference 86.

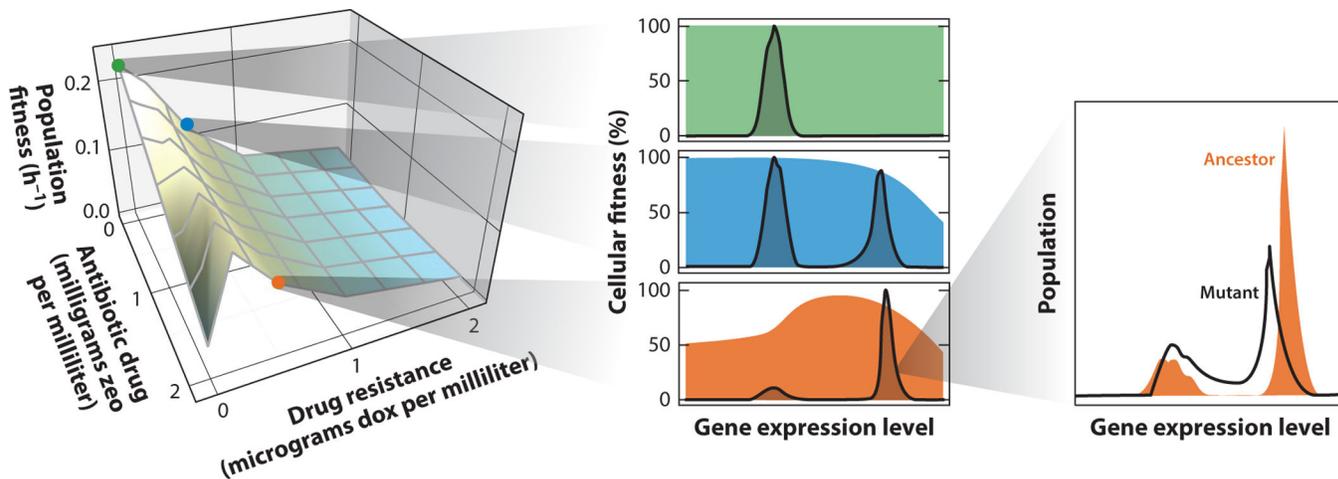


**Figure 12.**

Experimental cell fitness landscapes in systematically controlled environments. (a) Producing excess protein (LacZ) reduces cellular fitness. The blue dots and line represent the fitness in the absence of the inducer isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG), and the purple dots and line correspond to the environment with 1 mM IPTG. The red dot is a control strain with a deleted *lacY* gene in the presence of IPTG (89). (b) Increasing the concentration of an antibiotic drug in the medium diminishes bacterial fitness nonlinearly. Drug-resistant cells are more tolerant, but they are also more sharply inhibited at high drug concentrations. Drug resistance is measured by the activity of the the chloramphenicol-resistance enzyme chloramphenicol acetyltransferase. Lines of different colors represent the shape of the fitness landscape at fixed values of drug resistance. Panel *b* adapted with permission from Reference 94. (c) In fluctuating environments, cells evolve across a fitness moat to reach higher fitness in those complex environments. The variable  $E_0$  is the expression level of an operon affecting the growth rate in the absence of IPTG; the colored dots represent different stages of adaptation to reach the optimal fitness, progressing from green, to blue, to gray, to red. Panel *c* adapted with permission from Reference 4.

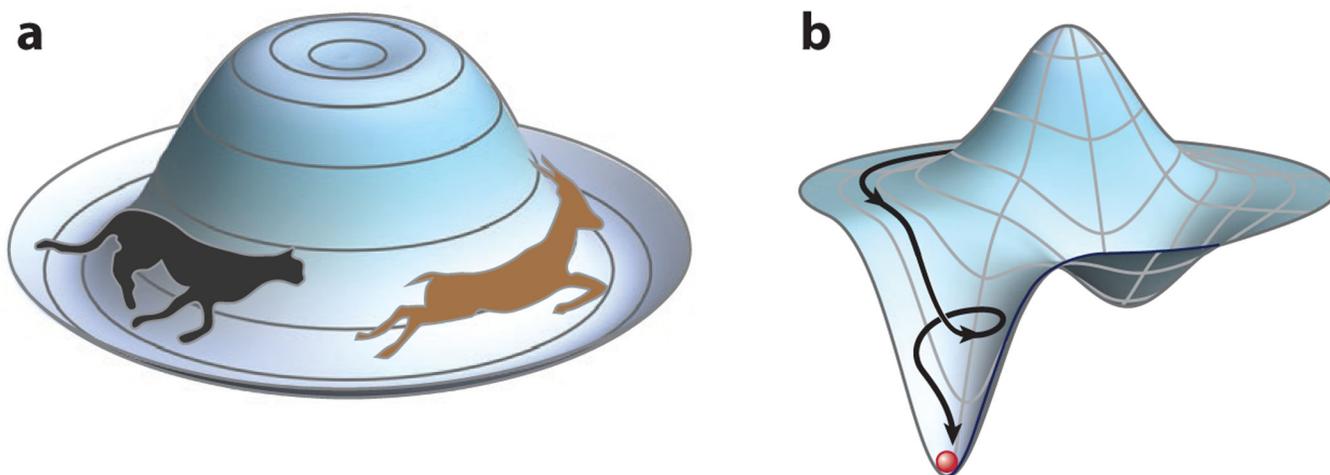


**Figure 13.** Population fitness landscape as a function of environmental factors. Each point on this landscape is the exponential growth rate of a yeast cell population for a given (antibiotic drug, drug resistance) combination. Experimentally measured values are indicated as colored dots. Abbreviations: atc, anhydrotetracycline; zeo, Zeocin.



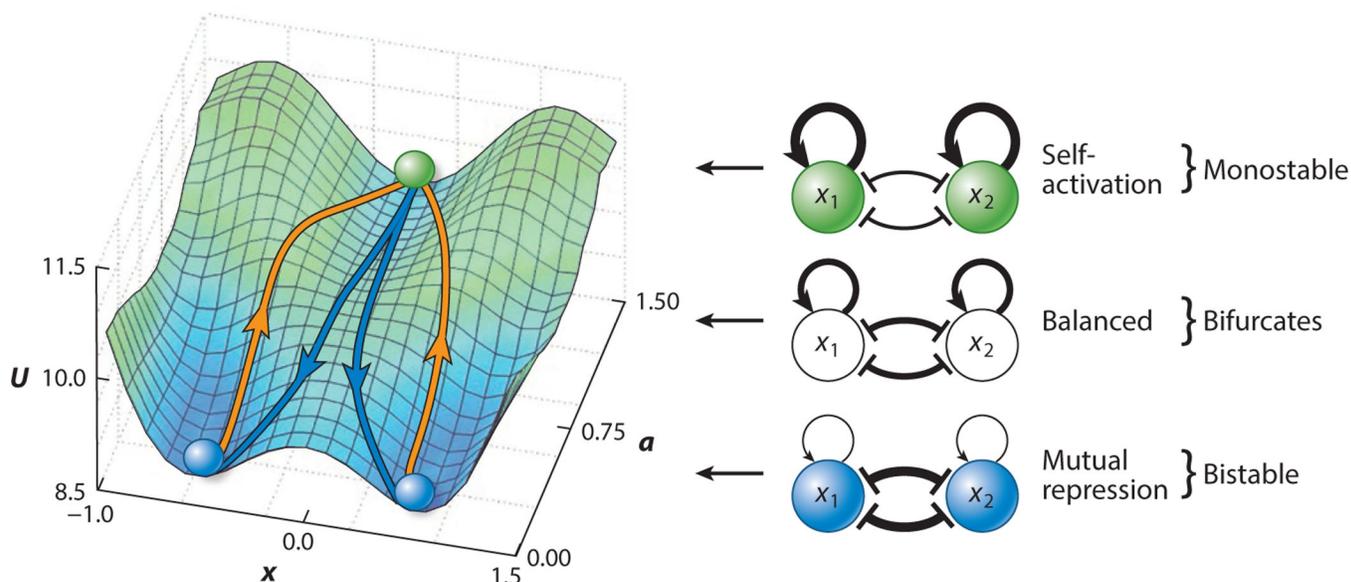
**Figure 14.**

The Goldilocks balance that leads to a just-right level of protein: not too much and not too little. Population fitness is the result of some microscopic factors: cellular fitness and gene expression. For a given cellular fitness landscape (*colored shading*), cellular gene expression (*black histograms*) can be either unimodal (overlaid with the *green fitness landscape*) or bimodal (overlaid with the *blue* and *orange fitness landscapes*). Each point on the macroscopic population fitness landscape (*colored dots*) results from weighted averaging of cellular fitness values over the corresponding gene expression distribution. Cellular fitness landscapes predict the evolution of gene expression changes: In constant environments, cells evolve toward unimodal gene expression located at the peak of each cellular fitness landscape. Abbreviations: dox, doxycycline; zeo, Zeocin.



**Figure 15.**

(a) Illustrating the Red Queen idea. Predator chases prey, and neither increases its own individual fitness in the evolutionary chase. (b) Illustrating the curl-flux principle. The population flow on a landscape is not directly down the gradient of a potential function; it also swirls if the system has an out-of-equilibrium driving force.



**Figure 16.**

The Waddington landscape and the new curl-flux-based understanding of it. The stem cell state is represented by the valley at the top of the landscape. The differentiated state is represented by the two valleys at the bottom. Metaphorically, Waddington differentiation is like a ball rolling from the top valley to the bottom ones. Two transcription factors can self-activate, other-repress, or degrade. The stem state is strong for self-activation (*top right*); the differentiated state is strong for repression (*bottom right*). What is new is our recent understanding of the curl-flux dynamics, showing how the stem cell state is also stable, why reprogramming is difficult, and how differentiation requires induction and is not just caused by spontaneous fluctuation (19). The full dynamical model shows that the stem cell state is stable, not unstable as in Waddington's static landscape. It shows that the reprogramming path is uphill. The dynamics model shows that the reprogramming path is not identical to the differentiation path. In the static model, the stem cell state (the hill) is not stable, so a small perturbation would allow spontaneous differentiation.