Napoleon Is in Equilibrium

Rob Phillips[1,2]

[1]Department of Applied Physics and Division of Biology, California Institute of Technology, Pasadena, California 91125; email: phillips@pboc.caltech.edu

[2]Laboratoire de Physico-Chimie Théorique, CNRS/UMR 7083-ESPCI, 75231 Paris Cedex 05, France

Annu. Rev. Condens. Matter Phys. 2015. 6:85–111

The *Annual Review of Condensed Matter Physics* is online at conmatphys.annualreviews.org

This article's doi:
10.1146/annurev-conmatphys-031214-014558

Keywords

allostery, MWC model, biophysics, gene regulation, transcription

Abstract

It has been said that the cell is the test tube of the twenty-first century. If so, the theoretical tools needed to quantitatively and predictively describe what goes on in such test tubes lag sorely behind the stunning experimental advances in biology seen in the decades since the molecular biology revolution began. Perhaps surprisingly, one of the theoretical tools that has been used with great success on problems ranging from how cells communicate with their environment and each other to the nature of the organization of proteins and lipids within the cell membrane is statistical mechanics. A knee-jerk reaction to the use of statistical mechanics in the description of cellular processes is that living organisms are so far from equilibrium that one has no business even thinking about it. But such reactions are probably too hasty given that there are many regimes in which, because of a separation of timescales, for example, such an approach can be a useful first step. In this article, we explore the power of statistical mechanical thinking in the biological setting, with special emphasis on cell signaling and regulation. We show how such models are used to make predictions and describe some recent experiments designed to test them. We also consider the limits of such models based on the relative timescales of the processes of interest.

"Everything flows, nothing stands still." – Heraclitus

1. EQUILIBRIUM IN A DYNAMIC WORLD

The mathematics that Newton invented in bringing us the modern conception of physics was fundamentally focused on change. Indeed, dynamics is one of the most cherished of topics in the physical sciences. But a strong case can be made that many of the most interesting examples of such change are those exhibited in the living world. Whether we watch time-lapse images of a growing vine as it encircles the branch of another plant, the synchronous and stereotyped embryonic development of hundreds of frog eggs engaged in their rhythmic progression toward a recognizable animal, or our own slow but unmistakable evolution revealed by a look in the mirror over the years as we see our aging frames, a sad reminder of Longfellow's musings,

> Art is long, and Time is fleeting,
> And our hearts, though stout and brave,
> Still, like muffled drums, are beating
> Funeral marches to the grave. . .,

it is clear even to the superficial observer that in life, all is change.

But how are we to come to terms with biological change? In this Annual Reviews series, Hopfield (1) gives us a fascinating account of his own first attempts at teaching biological physics and the question of how to handle the clear dynamism of living systems that goes well beyond the description offered here. The perhaps surprising tenet of this article is that even in the complicated, busy world of life, there are many circumstances in which we can make much progress, all the while invoking what at first blush look like strictly equilibrium notions. A colleague tells of hearing a seminar that invoked simple ideas from statistical mechanics to try and understand the distributions of regulatory proteins across genomic DNA. A noted biologist barked out "Napoleon is in equilibrium," the message being that ideas from equilibrium thermodynamics and statistical mechanics have no place in the study of living matter. There are problems with such a reaction on several fronts, namely, (*a*) that everyone does it; nearly every time we see histograms of protein binding occupancies on DNA from sequencing data there is a hidden occupancy assumption in play and at the very least, it is worth determining the limits and validity of the equilibrium occupancy null hypothesis and (*b*) depending upon the relative timescales of the processes of interest, sometimes equilibrium (or steady-state) ideas are demonstrably a defensible and useful quantitative approach. The argument presented in this review is that the subject is more nuanced than the apparently damning remark made by the noted life scientist and that the sometimes unreasonable effectiveness of equilibrium ideas makes for an interesting subject in its own right.

1.1. The Cell as the Test Tube of the Twenty-First Century

Test tubes conjure images of remaking nature rather than taking it as we find it. In test tubes, we mix together the components we have hypothesized will interact in some interesting way, we tailor the environment (for example, the temperature and pressure) to suit our needs, and we measure the time evolution of the reactants and products as they move toward their terminal privileged state of equilibrium. But how might we take the cells of living organisms and remake them in such a controlled way? Implementation of this program of remaking living cells had to await the genomic revolution of modern molecular biology, a revolution that made it possible to not only read

the information content of genomes at will but, even more impressively, to rewrite it to our own purposes (2). It is now the province of the field of synthetic biology to remake cells, reprogramming their genes so that cells light up with fluorescence where and when we want, knocking out their genes one by one at will, and rewiring their genetic networks to make switches and oscillators of all kinds from scratch (3–6).

The way synthetic biology can drive physical biology has many parallels with the way that materials science has fueled advances in condensed matter physics through the production of new classes of materials. Using the tools of genetic engineering, whole new classes of living materials can be produced, making it possible to ask new and precise questions about biological phenomena in organisms ranging from the simplest of bacteria to the yeast cells that make some of our delicious foods and beverages possible to the flies that circle over those foods in our kitchens. Indeed, freakish mutant flies have been synthesized Frankenstein-like with photoreceptors from the eyes growing on parts of the body where they would never belong (7). But this is more than just parlor tricks. By rewiring cells to exploit a technique known as optogenetics, in which light is used to control the state of activity of a given molecule, it is now possible to drive cellular decision making with nearly arbitrary temporal histories (8–10). By rewiring biology, we understand it in ways that are not possible if we simply take it as evolution has left it, effectively delivering on the promise of turning cells (and even entire multicellular organisms) into test tubes.

One of the reasons that the question of how to model the dynamics of the many molecular processes that animate cells is so timely is that the experimental state of the art has reached the point at which, thanks to the synthetic biology revolution, reproducible, quantitative measurements are becoming routine. For example, studies on cell signaling have reached the point at which the response function of cells subjected to time-varying perturbations, such as osmotic stresses or chemoattractants, can be measured directly (11–15). Using microfluidics in conjunction with fluorescence microscopy, it is now possible to measure the time history of a variety of molecular readouts to perturbations, such as temporal changes, in the presence of chemoattractant or time-varying alterations in the osmolarity of the cellular medium. Another class of input-output functions that has been characterized in great detail are those associated with the processes of the central dogma of molecular biology (see **Figure 1**), namely, transcription (i.e., the production of mRNA molecules) and translation (i.e., the synthesis of new proteins), as they occur throughout the cell cycle (17–21). It is the theoretical description of the dynamics of the central dogma that forms the centerpiece of this article, with special reference to the process of transcription.

In my view, one of the most important insights to emerge from the quantitative study of the physical world is the ability to construct effective theories that make no explicit reference to the fundamental constituents at smaller scales. Atomic physicists treat the nucleus as a point particle with no internal degrees of freedom. More blatantly, elastic theories of solid materials and hydrodynamic theories of fluids ignore molecular reality altogether, although the material parameters, such as elastic moduli or viscosities, that show up in such theories can be computed by appealing to this molecular reality. When first learning about these continuum theories, we are often instructed to think of a control volume that is large on the scale of individual atoms or molecules and to imagine states of deformation within materials that are long wavelength with respect to atomic dimensions (22). But interestingly, these kinds of descriptions often work on much smaller scales than one might expect strictly on the basis of the control-volume picture (23).

To my mind, one of the most important scientific challenges we face in coming to terms with the sometimes dazzling complexity of biological systems is figuring out just how far we can go down this same path of constructing effective theories that self-consciously root

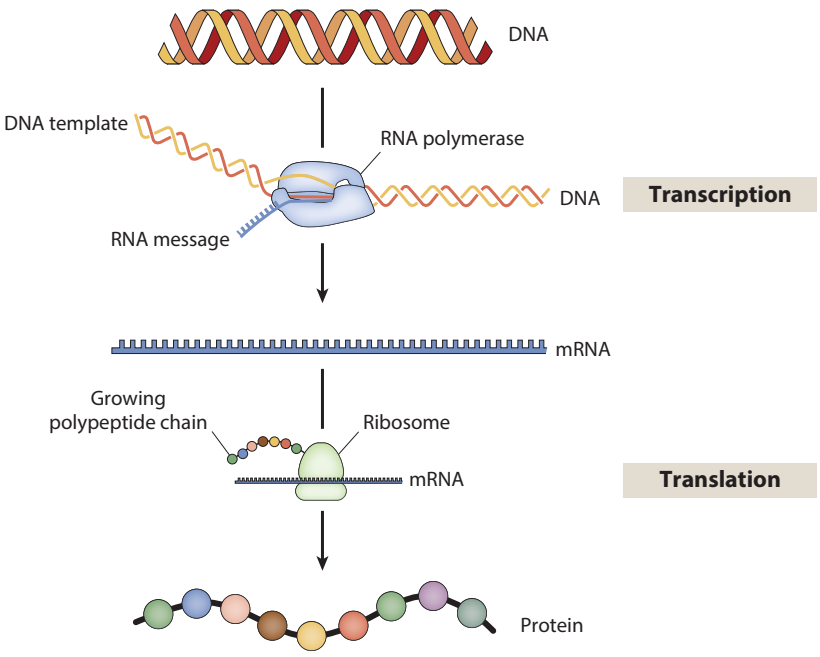


Figure 1

The pathway from DNA to protein. When a gene is on, the protein molecular copying machine known as RNA polymerase translocates along that gene, producing an mRNA template of that gene. This mRNA template is then read by the molecular machine known as the ribosome, which produces a polypeptide chain by stringing together amino acids in accordance with the triplets of codons specified by the mRNA. Adapted from Reference 16.

out some features of the problem. Such theories are sometimes viewed with suspicion in biology, passingly referred to as just phenomenological, and hence the challenge is to show how they can be used to generate falsifiable predictions that could not have been arrived at from the first-principles perspective in much the same way that the beautiful patterns, such as convection cells seen in clouds, are not usually described on the basis of molecular dynamics.

In the remainder of this article, we explore the limits and validity of equilibrium thinking in informing one particular set of questions in cell biology: What are the mechanisms whereby cells make decisions? However, the same strategy advocated here should be viewed as being relevant in a host of other contexts, ranging from the rules governing the opening and closing of ion channels in response to various environmental stimuli to the induction of signaling cascades when bacteria detect chemoattractants in their environment to the reprogramming of cells between different cell fates. Statistical mechanics' great promise was as a way to deal with complicated systems of many particles. How far can we take those ideas in thinking about the messy and complicated network of molecular interactions within a cell?

Note also that this article is much more about providing a point of view than it is about surveying a vast and deep literature. The works I have cited have been chosen as representative and not with the ambition of assigning priority and recounting the history of this enormous field. I am fully cognizant of the many important contributions that have not been included and encourage interested readers to use the references provided here as a jumping-off point for a more rigorous examination of the vast and interesting literature.

2. GENE REGULATION AND STATISTICAL PHYSICS

2.1. How Cells Decide

One fertile arena for exploring the question of how to come to quantitative and predictive terms with the complexity of biological systems centers on perhaps the most unique feature of living matter, namely, the genome. As every high school kid now learns, genetic information is stored in the enormous DNA molecules making up the chromosomes of a given organism. These chromosomes, in turn, are partitioned into collections of bases that are the molecular incarnation of the abstract particles of inheritance introduced by Mendel and are known as genes, with each gene starting with a so-called start codon (ATG) and ending with one of several possible stop codons. In a simple bacterium such as *Escherichia coli*, which has a genome of roughly 5×10^6 base pairs, we can estimate the number of genes by noting that a typical protein in a bacterium has a length of roughly 300 amino acids and that it takes three nucleotides (i.e., A, T, G and C) to code for each such amino acid through the famed genetic code. This means that roughly 1,000 base pairs are used to code for each protein and hence that the 5×10^6 base pairs of the genome can be used to code for roughly 5,000 contiguous genes.

But how are these genes actually accessed and turned into physiological responses? This takes place through the expression of those genes in the process of transcription, in which RNA polymerase (RNAP) [$\approx 10^3 - 10^4$ copies per cell (24)] reads the sequence of nucleotides on the DNA molecule and synthesizes a corresponding mRNA molecule as highlighted in the top part of **Figure 1**. Indeed, one of the signature achievements of modern molecular biology (and I would go further and say in the entire history of science) was the determination of the steps, often known as the central dogma of molecular biology, leading from DNA to active proteins. However, beautiful experiments of many different types, whether in single-celled bacteria or the multicellular collection known as a human being, tell us that under normal circumstances only a limited battery of these genes are actually on and being used to synthesize new proteins. The transcription process in bacteria occurs after RNAP binds to a region in front of the gene of interest known as a promoter.

One of the ways control is exercised over the transcription of a given gene is via a process known as transcriptional regulation. In this process, batteries of proteins known as transcription factors bind to the DNA and either tune down (repression) or tune up (activation) the level of expression of some gene of interest. These transcription factors are, therefore, themselves the products of gene expression, implying that there are interesting networks featuring feedback between protein synthesis and regulation. The question of how the decision about which genes are on is made is of critical importance in a broad array of topics, ranging from the development of an egg to an embryo to the unchecked proliferation of cells in the growth of a tumor. We use this question of cellular decision making as our entry into the question of how statistical mechanics can be used in the service of answering biological questions.

2.2. Thermodynamic Models of Gene Regulation

Our starting point for thinking quantitatively about the transcription process is to consider the kinetics of mRNA production. In particular, we are interested in the time evolution of the mRNA census as a result of the key dynamical processes, such as transcription itself and mRNA degradation. Specifically, we imagine a collection of N cells as shown in **Figure 2** at time t in which each cell harbors a certain number of mRNA molecules. This distribution of mRNA molecules throughout our collection of cells can be represented by the histogram shown at the top of **Figure 2**.

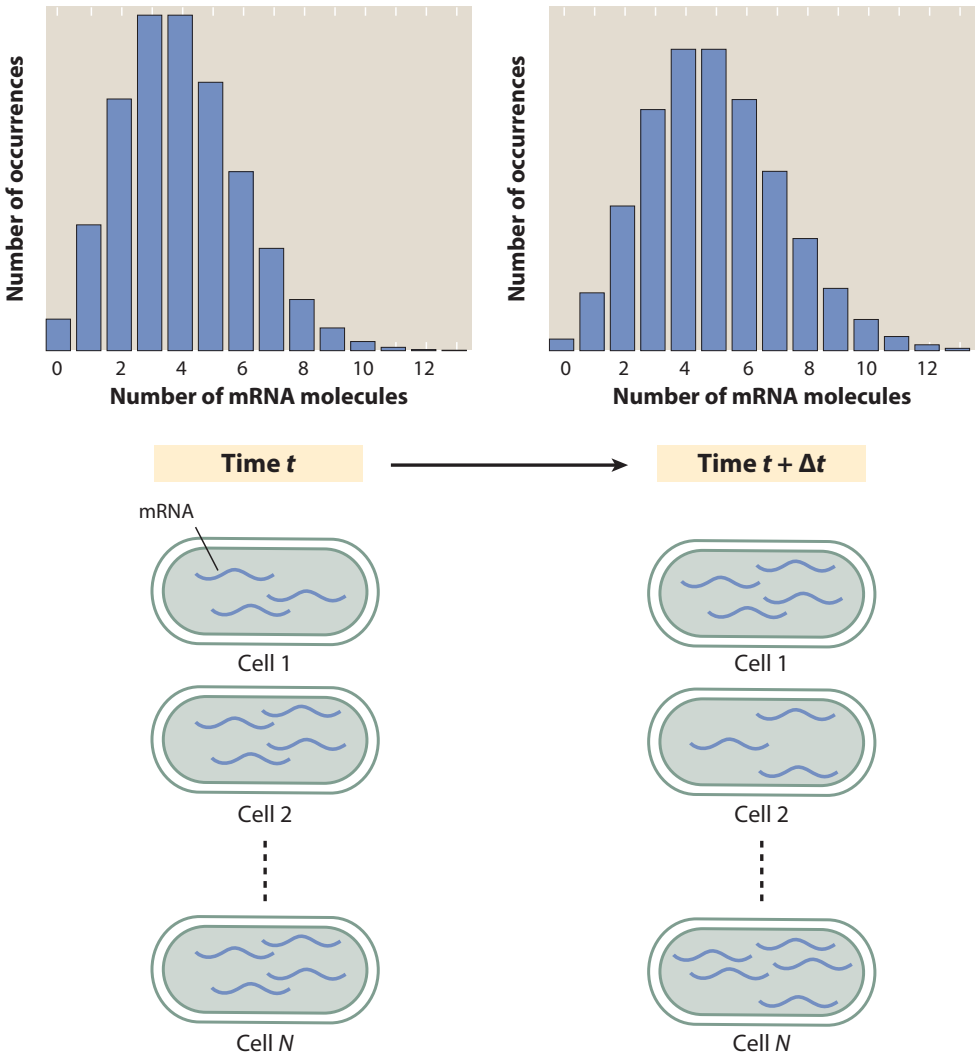


Figure 2

Transcription process resulting in change in mRNA census between times t and $t + \Delta t$. The schematic histogram shows the distribution of the number of mRNA molecules found per cell. We refer to the average number of mRNA at time t as $m(t)$; it is found by adding up the total number of mRNA over all cells and dividing by the number of cells. The number of mRNA per cell increases because of transcription and decreases because of mRNA degradation.

By summing up all of the mRNA molecules in all of the cells $M = \sum_i m_i$, we can now find the average number of mRNA molecules per cell as

$$\langle m \rangle = \frac{M}{N}, \qquad 1.$$

which, for simplicity of notation, we label as $m(t)$ without the $\langle m \rangle$ symbols that make the averaging process explicit.

The time evolution of the probability distribution $p(m, t)$ can be worked out using a chemical master equation that acknowledges these different transcriptionally active states (25–30). Such equations, in turn, can be used to generate dynamical equations for the mean number of mRNA per cell. Given that the average amount of mRNA per cell at time t is $m(t)$, the amount a small time increment later, Δt, is as seen in **Figure 2** given by

$$m(t + \Delta t) = m(t) - \text{number of mRNA degraded} + \text{number of mRNA produced}. \quad 2.$$

To make further progress, we must be able to write the degradation and production terms explicitly. To that end, we exploit the fact that in the simplest model of degradation, at every instant each mRNA molecule has the same probability $\gamma\Delta t$ of decaying. For a gene that is under no transcriptional control (i.e., a so-called constitutive promoter), mRNA is produced at a steady rate (r). For genes subjected to transcriptional control, there can be a variety of transcriptionally active states. For example, the state in which just RNAP is bound corresponds to one transcriptionally active state, whereas a state with both polymerase and an activator molecule bound simultaneously can have an even higher rate of transcription. For each of the transcriptionally active states, there is a rate, r_i, of mRNA production, as shown in **Figure 3**. Given these ideas, the amount of mRNA at time $t + \Delta t$ can be written more formally as

$$m(t + \Delta t) = m(t) - \gamma\Delta t m(t) + \sum_i (r_i \Delta t) p_i, \quad 3.$$

where γ is the degradation rate and has units of time^{-1}, and p_i is the probability of being in the ith transcriptionally active state.

But what is the nature of the individual states and their probabilities? Although their explicit calculation is the business of the thermodynamic models that are our focus below, here we note that the probability of the ith transcriptionally active state can be thought of as

$$p_i = p_i([TF_1], [TF_2], \cdots), \quad 4.$$

where the notation indicates that this probability is a function that reflects the occupancy of the regulatory DNA by the various transcription factors (i.e., regulatory proteins) that interact with the regulatory apparatus of the gene of interest. Hence, each transcriptionally active state denoted by the label i corresponds to a different state of the promoter characterized by a different constellation of bound transcription factors. These ideas were first put into play carefully in the regulatory setting by Ackers (31, 32) and coworkers and have been explored more deeply in the meantime by a number of groups (33–37).

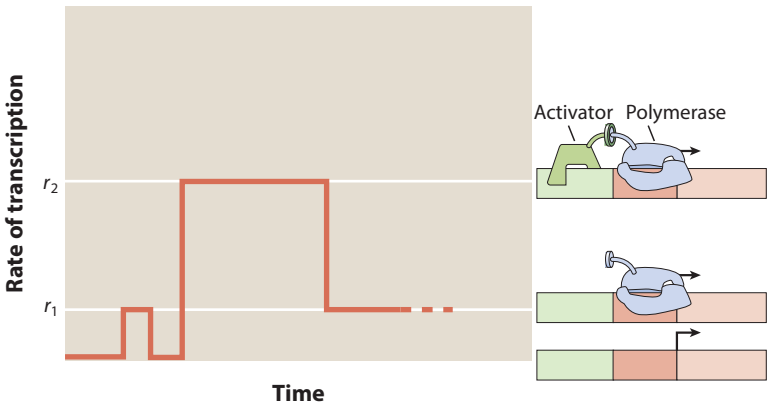


Figure 3

Transcription rate in different states. This schematic shows the time evolution of the transcription rate of the gene of interest as it transitions between different transcriptionally active states. The icons on the right show that when RNA polymerase is absent, the rate is 0; when just polymerase is present, the rate is r_1; and when polymerase and activator are both present, the rate is r_2.

It is more traditional to rewrite our equation for mRNA time evolution in the form of a differential equation as

$$\frac{dm}{dt} = -\gamma m + \sum_i r_i p_i, \qquad 5.$$

where, as above, p_i is the probability of the ith transcriptionally active state, and r_i is the rate of transcription when the system is in that state. The individual p_is are computed based on the occupancy of various constellations of transcription factors in the promoter region of the gene of interest.

The simplest example of the kinds of models given above is the constitutive promoter in which there are no regulatory interventions by transcription factors. In this case, there are only two states: the state in which the promoter is empty and thus transcriptionally silent and the state in which the promoter is occupied by RNAP and hence transcriptionally active with rate r. We can write the dynamical equation for mRNA production in this case as

$$\frac{dm}{dt} = -\gamma m + r p_{bound}, \qquad 6.$$

where we can make contact with Equation 5 by noting that there is only one transcriptionally active state such that $r_1 = 0$ and $r_2 = r$, and we note that the probability of that transcriptionally active state is $p_2 = p_{bound}$. The central preoccupation of the thermodynamic class of models is how to determine the quantity p_{bound}, which in the simplest version of these models refers to the probability that RNAP is bound to our promoter of interest. But how do we compute this probability using the tools of statistical mechanics?

Figure 4 gives a schematic view of the kinds of abstractions and simplifications we make in thinking about the binding of proteins, such as RNAP and transcription factors, to the genome. In particular, we imagine the genome as a linear array of N_{NS} boxes (equal to the number of base pairs in the genome), each of which is a possible binding site for the RNAP and for transcription factors. Once we have identified the different states of the promoter of interest, our next task is to assign statistical weights to those separate states, which is a prerequisite to our being able to write the probabilities of those states themselves. To compute the statistical weights, we need to evaluate the energies of the different states as shown in **Figure 5**. In the simplest of models, we imagine that nonspecifically bound polymerases have energy $\varepsilon_{pd}^{NS} < 0$ and can be bound to N_{NS} distinct sites (i.e., number of base pairs) on the genome. When a polymerase is bound to its promoter, the corresponding energy is $\varepsilon_{pd}^{S} < 0$. Note that these energies are effective parameters, as discussed earlier in the paper, with the specific binding energy more energetically favorable than the nonspecific binding energy (i.e., $\varepsilon_{pd}^{S} < \varepsilon_{pd}^{NS}$). We have no first-principles access to these numbers because they depend upon a host of factors, including the salt concentration, hydrophobic interactions, van der Waals interactions, specific contacts between residues on the proteins and bases in the DNA, etc. Hence, in much the same way that in the Ising model we lump our ignorance of microscopics into effective coupling parameters, here the messy business of protein-DNA interactions is subsumed in several effective parameters.

With these energies in hand, Boltzmann tells us that the way to find the statistical weight w_i of some set of configurations all having the same energy is

$$w_i = g_i(E_i) e^{-\beta E_i}, \qquad 7.$$

where E_i is the energy of that state and $g_i(E_i)$ tells us how many states there are that have that energy. The quantity $g_i(E_i)$ is the multiplicity of the state with energy E_i and can be computed using combinatorial ideas like those indicated schematically in **Figure 4*c***.

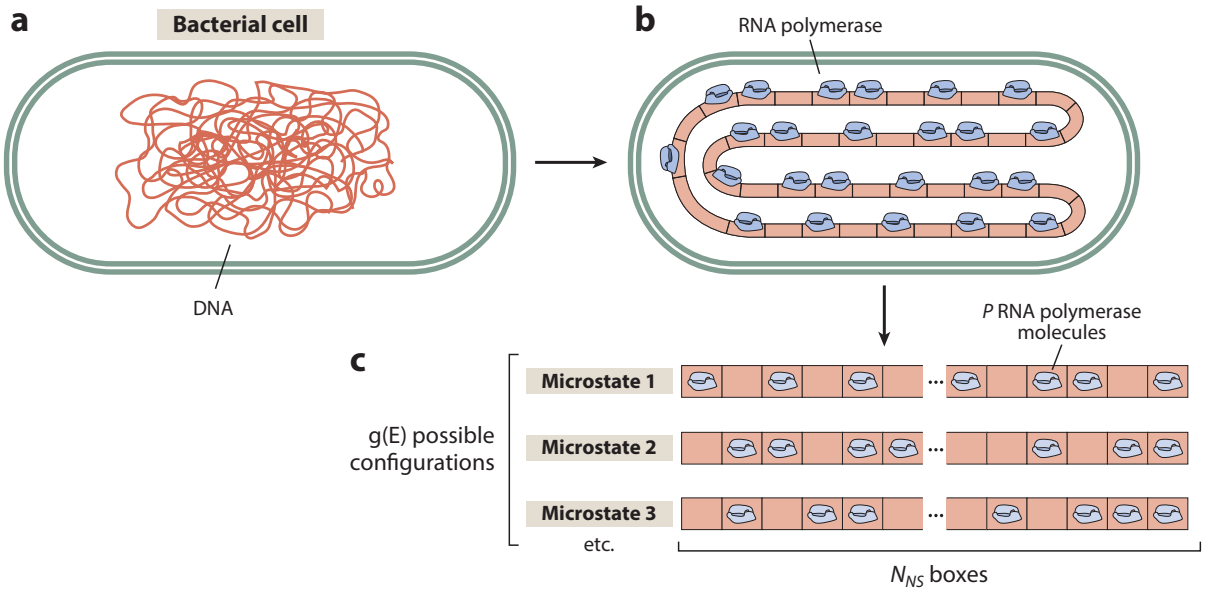


Figure 4

The Boltzmann genome. (*a*) Schematic of a bacterial cell showing the complicated internal arrangement of the genome. (*b*) Abstraction of the genomic landscape of a bacterium into a one-dimensional lattice of binding sites. (*c*) There are many microstates [g(E) of them] of the *P* RNA polymerase molecules on the genomic DNA, and this can be evaluated combinatorially by thinking of the genome as consisting of N_{NS} nonspecific binding sites upon which the *P* RNA polymerases can be arranged.

We can implement the statistical mechanical program of determining the partition function for a constitutive (unregulated) promoter, as shown in **Figure 5**. Given the statistical weights of the two different states (i.e., unoccupied promoter and promoter occupied by RNAP), the probability of the transcriptionally active state is given as the ratio

$$p_{bound} = \frac{\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}{1 + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}, \tag{8.}$$

where $\Delta \varepsilon_{pd} = \varepsilon_{pd}^{S} - \varepsilon_{pd}^{NS}$. Essentially, this is nothing more than a traditional noncooperative binding curve. As the number of RNAP molecules increases, the probability of binding saturates.

For the case of a promoter that is regulated by a single transcription factor that represses that gene, the various categories of states of the system are shown in **Figure 6**. Using precisely the same reasoning, we can compute the probability that the promoter is occupied as

$$p_{bound} = \frac{\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}{1 + \frac{R}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd}} + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}. \tag{9.}$$

The new parameters that enter this expression are the number of repressors (R) and the difference between specific and nonspecific binding of repressors given by $\Delta \varepsilon_{rd}$. If we introduce the simplifying notations $p = (P/N_{NS}) e^{-\beta \Delta \varepsilon_{pd}}$ and $r_A = (R/N_{NS}) e^{-\beta \Delta \varepsilon_{rd}}$, then we can rewrite our expression for the gene being in the on state as

$$p_{bound} = \frac{p}{1 + r_A + p}. \tag{10.}$$

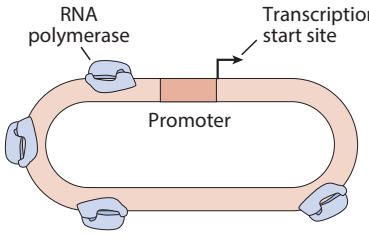
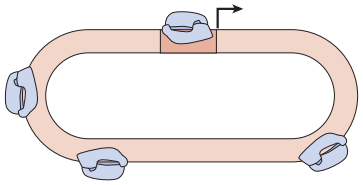
State	Energy	Multiplicity	Weight (Multiplicity × Boltzmann weight)	Rate
	$P\varepsilon_{pd}^{NS}$	$\frac{N_{NS}!}{P!\,(N_{NS}-P)!} \approx \frac{(N_{NS})^P}{P!}$	$\frac{(N_{NS})^P}{P!}\, e^{-P\varepsilon_{pd}^{NS}/k_BT}$	0
	$(P-1)\varepsilon_{pd}^{NS}+\varepsilon_{pd}^{S}$	$\frac{N_{NS}!}{(P-1)!\,[N_{NS}-(P-1)]!} \approx \frac{(N_{NS})^{P-1}}{(P-1)!}$	$\frac{(N_{NS})^{P-1}}{(P-1)!}\, e^{-(P-1)\varepsilon_{pd}^{NS}/k_BT}\, e^{-\varepsilon_{pd}^{S}/k_BT}$	r

Figure 5

States and weights of the constitutive (unregulated) promoter. The transcription start site is shown by a bent arrow. The first state is transcriptionally inactive and has the transcriptional rate $r_1 = 0$, and the second state is transcriptionally active and is assigned the transcriptional rate $r_2 = r$.

From the perspective of experimental measurements on transcription, often it is most convenient to characterize the level of expression by a comparative metric known as the fold-change. The idea of the fold-change is to compare how much expression there is in the presence of the regulatory machinery to that in its absence. Specifically, this ratio can be written as

$$\text{fold-change} = \frac{p_{bound}(R \neq 0)}{p_{bound}(R = 0)}. \qquad 11.$$

For the simple repression architecture considered here, the fold-change simplifies to the form

$$\text{fold-change} = (1 + r_A)^{-1}, \qquad 12.$$

where we have invoked the so-called weak-promoter approximation, in which $p \ll 1$. What this tells us is that as the number of repressors gets larger (R increases), the fold-change becomes smaller and smaller.

In order for theoretical predictions like those described above to have real meaning, there must be a concomitant effort to construct experiments designed to explicitly test the scaling results implied by those theories. Two of the key criteria that must be met by such measurements are precision and reproducibility. The best treatment of the subject of reproducibility that I am aware of is now a century old and emerged in the context of early studies on the existence of atoms. In his classic treatise on the topic, French physicist Jean Baptiste Perrin presents a stunning table that shows 15 distinct ways of determining Avogadro's number (38). Again, to our way of thinking, these are the kinds of standards that biological research needs to be held to in order for the kind of theory-experiment interplay enjoyed by condensed matter physics to take root in the study of living matter.

Detailed experimental exploration of the response of the simple repression architecture (i.e., Equation 12) is shown in **Figure 7**, where we see that precisely the kind of scaling with the repressor number expected from the thermodynamic framework is observed experimentally,

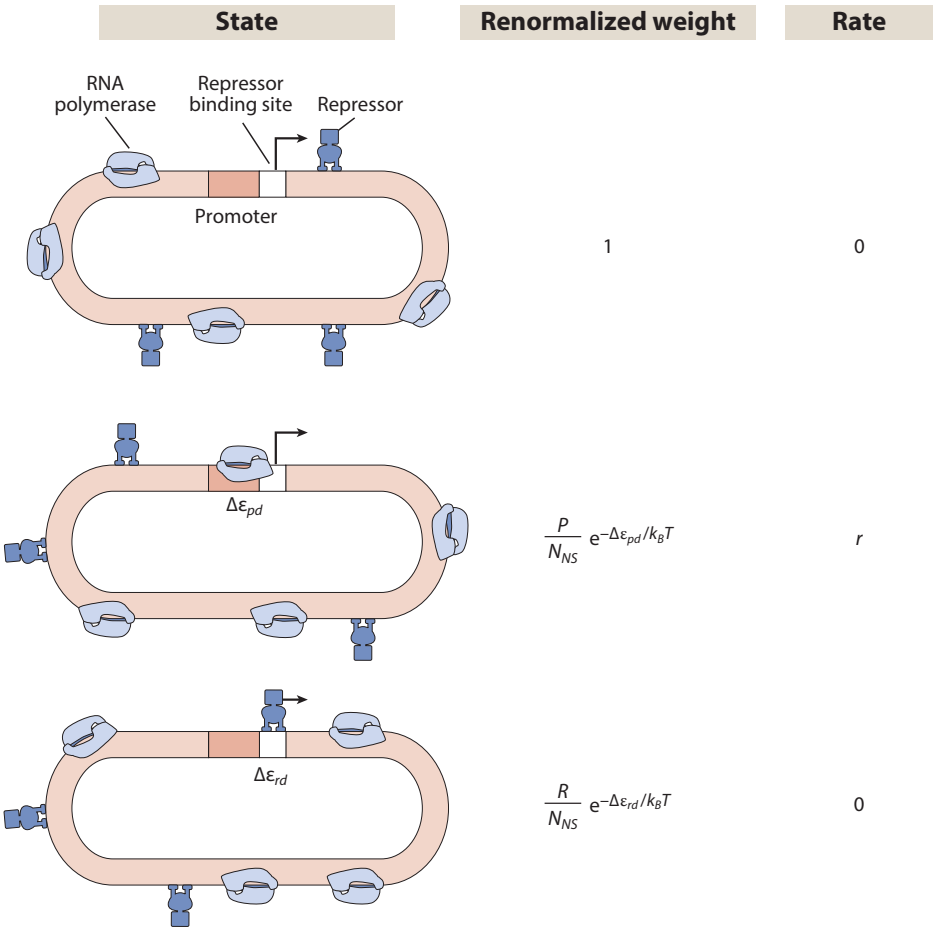


Figure 6

States and weights for the simple repression regulatory architecture. The binding site for the repressor molecule is shown as a white fragment of the genomic DNA. The first state shows the empty promoter, the second state shows the transcriptionally active (i.e., polymerase-bound) state, and the third state is the repressed state in which the promoter is rendered inaccessible to polymerase binding as a result of repressor binding.

although here the readout of gene expression was the protein products of the mRNA rather than the mRNA itself. Given that the quantity r_A in Equation 12 is given by $r_A = (R/N_{NS})e^{-\beta\Delta\varepsilon_{rd}}$, this implies that differences in fold-change can be elicited not only by tuning the repressor number but also by changing the strength of the repressor binding sites as reflected in the parameter $\Delta\varepsilon_{rd}$. In particular, by changing the sequence of the repressor binding region on the DNA, a roughly 20-bp fragment, we can tune the parameter $\Delta\varepsilon_{rd}$. **Figure 7** shows four distinct choices of the repressor-DNA binding energy with stronger repressor-DNA binding corresponding to a more substantial fold-change in gene expression. Note that the comparison between theory and experiment revealed in the single-cell measurements reported in **Figure 7** involves no free parameters. Specifically, the repressor number is measured directly, and the binding energies $\Delta\varepsilon_{rd}$ were determined from a single measurement at one repressor copy number in the bulk setting, meaning that in the context of the single-cell measurements shown in the figure, the comparison between theory and experiment required nothing more than simply plotting the theoretical result of Equation 12 and the experimental results on the same figure.

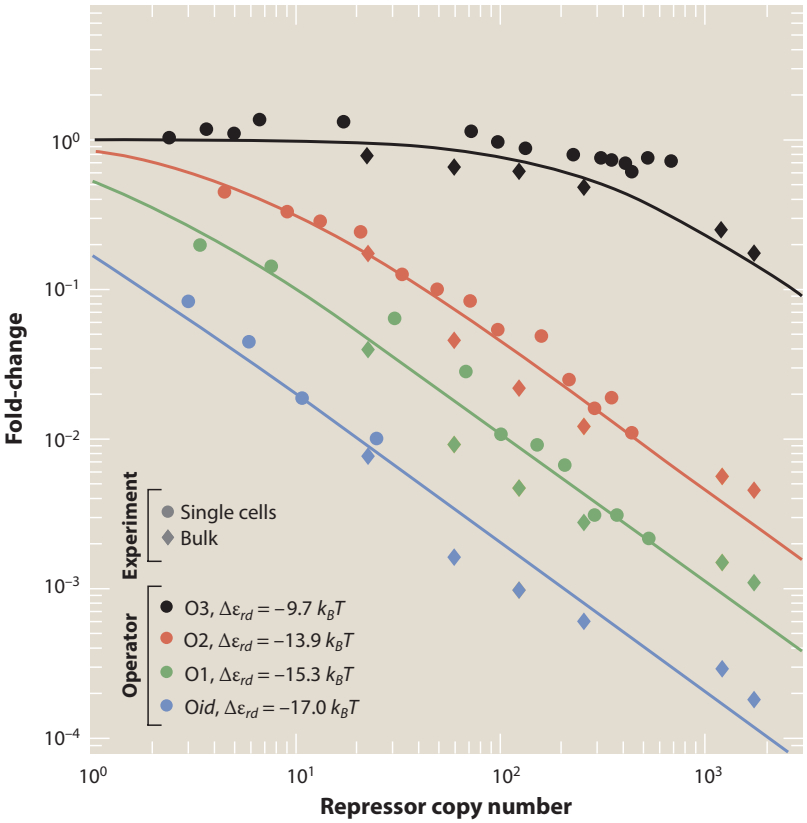


Figure 7

Fold-change in the simple repression architecture. The y-axis shows how much less expression of the gene of interest we get in the presence of repressors than in their absence. As the number of repressors increases, the expression of the gene of interest is reduced even more. The different curves correspond to different binding site strengths for the repressor on its genomic binding site. From top to bottom, the highest curve shows the response for the weakest binding site, and the lowest curve shows the response for the strongest binding site. The circles show the result of measuring the gene expression and repressor number using fluorescence at the single-cell level, and the diamond shapes show the results from bulk assays in which the number of repressors is counted using quantitative immunoblots and the level of expression is read out through the action of an enzyme produced when the gene is on (39, 40).

2.3. Uses and Abuses of Occupancy

Interestingly, there seems to be anecdotal agreement that when writing dynamical descriptions of the processes of the central dogma, such as transcription, the dynamics can be described by occupancy functions, such as those presented in Equation 5. That is, in order to figure out the level of transcription, we figure out what fraction of the time the promoter is present in each of its transcriptionally active states and sum up the contributions from each of those states. However, this is a hypothesis that needs to be scrutinized relentlessly. Specifically, the occupancy hypothesis should be used to generate polarizing predictions and then experiments should be designed to put those predictions to the test. Recent single-cell experiments have begun to do just that (41, 42).

In the previous section, we showed how by treating the occupancy from the perspective of the Boltzmann equilibrium probabilities, we could erect the entire machinery of thermodynamic models. An alternative way of viewing these same problems that begins to address the "Napoleon is in equilibrium" objection is by appealing to the kinetic theory of transcription

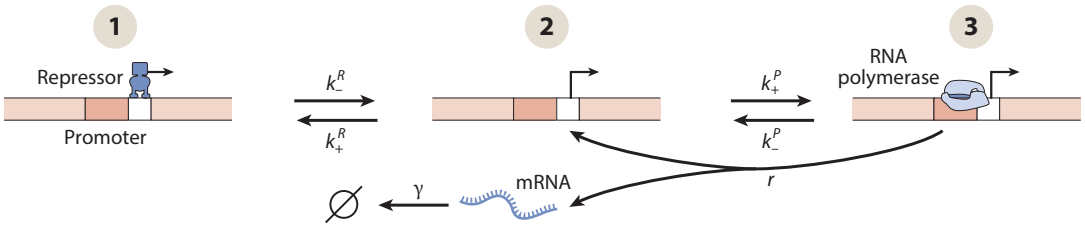


Figure 8

Effective model of kinetics of mRNA production for the simple repression regulatory motif. The promoter transitions between the empty state (state 2), the repressed state (state 1), and the transcriptionally active state (state 3), with rate constants denoted in the figure. When in state 3, transcription can begin, which leads to the production of mRNA and the return of the promoter to state 2.

and working out the time evolution of the probabilities of the different states (25–30, 43, 44). Here, to choose a specific regulatory architecture as an example, we think of the simple repression motif as a chemical reaction of the form shown in **Figure 8**. mRNA is produced at rate r when the system is in state 3. The three states correspond to precisely the three states shown in **Figure 6**, although note that in this figure the ordering of states has been changed. The rate constants shown in the reaction scheme dictate the frequency of switching between the different states.

We begin with working out the kinetics of the first state, which has the promoter occupied by the repressor molecule. As the linear reaction scheme shows, there is only one way to enter and exit this state and that is through the empty state (state 2), resulting in the dynamical equation

$$\frac{dp(1)}{dt} = k_+^R p(2) - k_-^R p(1). \qquad 13.$$

The dynamics of the empty state (state 2) are more complicated because this state is accessible to both the repressor and the polymerase, meaning that the dynamics can be written as

$$\frac{dp(2)}{dt} = -k_+^R p(2) + k_-^R p(1) - k_+^P p(2) + k_-^P p(3) + rp(3). \qquad 14.$$

Note that the final term in this equation reflects the fact that mRNA is produced at rate r from state 3 and once mRNA production begins, polymerase leaves the promoter and hence the system goes back to state 2. The state with polymerase occupying the promoter evolves similarly as can be seen by writing

$$\frac{dp(3)}{dt} = k_+^P p(2) - k_-^P p(3) - rp(3). \qquad 15.$$

To close the loop and come full circle to the real question of interest, namely, the production of mRNA itself, we have

$$\frac{dm}{dt} = -\gamma m + rp(3). \qquad 16.$$

What this equation tells us is that the promoter is only transcriptionally active in the third state, namely, that state in which the polymerase binds the promoter.

These equations can be resolved in steady state by setting the left side in each case to zero. We begin by solving for the steady-state level of mRNA and find

$$m_{ss} = \frac{rp(3)}{\gamma}. \tag{17.}$$

But what is $p(3)$? In seeking the unknown steady-state probabilities, we must respect the constraint that the probabilities sum to one, namely,

$$p(1) + p(2) + p(3) = 1. \tag{18.}$$

We do not go into the details of the algebra of resolving these three linear equations and simply quote the results as

$$p(1) = \frac{\frac{k_+^R}{k_-^R}\frac{(k_-^P + r)}{k_+^P}}{1 + \frac{(k_-^P + r)}{k_+^P}\left(1 + \frac{k_+^R}{k_-^R}\right)}, \tag{19.}$$

$$p(2) = \frac{\frac{(k_-^P + r)}{k_+^P}}{1 + \frac{(k_-^P + r)}{k_+^P}\left(1 + \frac{k_+^R}{k_-^R}\right)}, \tag{20.}$$

and

$$p(3) = \frac{1}{1 + \frac{(k_-^P + r)}{k_+^P}\left(1 + \frac{k_+^R}{k_-^R}\right)}. \tag{21.}$$

Note that with these steady-state probabilities in hand, we can in turn determine the steady-state mRNA number. The probabilities resulting from the kinetic model are contrasted with those from the thermodynamic model in **Figure 9**.

To make contact with the kinds of experiments discussed earlier and presented in **Figure 7**, we have to compute the fold-change given as

$$\text{fold-change} = \frac{m_{ss}(R \neq 0)}{m_{ss}(R = 0)} = \frac{1}{1 + \frac{\frac{(k_-^P + r)}{k_+^P}}{1 + \frac{(k_-^P + r)}{k_+^P}}\frac{k_+^R}{k_-^R}}. \tag{22.}$$

Note that we can write $k_+^R = k_+ R$, where we have acknowledged that the on rate for the repressor is proportional to how many repressors are present in the cell. Interestingly, we see that this implies that the functional form of the fold-change is the same even in this steady-state context, as it was in the thermodynamic model framework, although now at the price of having to introduce an effective K_d^{eff}, resulting in

$$\text{fold-change} = \frac{1}{\left(1 + \frac{R}{K_d^{eff}}\right)}. \tag{23.}$$

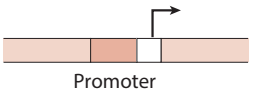
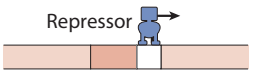
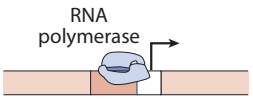
State	Steady-state probability	Boltzmann probability	Rate of promoter escape
	$\frac{1}{1+\frac{k_-^P+r}{k_+^P}\left(1+\frac{k_+^R}{k_-^R}\right)}$	$\frac{1}{1+\frac{P}{K_d^P}+\frac{R}{K_d^R}}$	0
	$\frac{\frac{k_+^R}{k_-^R}\frac{k_-^P+r}{k_+^P}}{1+\frac{k_-^P+r}{k_+^P}\left(1+\frac{k_+^R}{k_-^R}\right)}$	$\frac{\frac{R}{K_d^R}}{1+\frac{P}{K_d^P}+\frac{R}{K_d^R}}$	0
	$\frac{\frac{k_-^P+r}{k_+^P}}{1+\frac{k_-^P+r}{k_+^P}\left(1+\frac{k_+^R}{k_-^R}\right)}$	$\frac{\frac{P}{K_d^P}}{1+\frac{P}{K_d^P}+\frac{R}{K_d^R}}$	r

Figure 9

Comparison of states and weights from kinetic and thermodynamic models.

By comparing Equations 12 and 23, we see that their scaling with the repressor number is identical. To further explore the common features between these two expressions for fold-change, note that

$$K_d^{eff} = \frac{k_-^R}{k_+}\frac{\left(1+\frac{(k_-^P+r)}{k_+^P}\right)}{\frac{(k_-^P+r)}{k_+^P}}. \qquad 24.$$

We can simplify this further by noting that we can write $K_d^{(R)} = k_-^R/k_+$, resulting in

$$K_d^{eff} = K_d^{(R)}\frac{\left(1+\frac{(k_-^P+r)}{k_+^P}\right)}{\frac{(k_-^P+r)}{k_+^P}}. \qquad 25.$$

One conclusion to emerge from this result is that the idea of testing the thermodynamic model, even in the case of simple repression, is more subtle than we might have thought given that correspondence with the functional form given here could be a signature of consistency either with the thermodynamic model or with the kinetic theory of transcription. However, by using experiments to probe the intrinsic dependence of K_d^{eff} on other quantities, such as how fast polymerases leave the promoter, and on the number of polymerases, we can begin to understand whether the apparent accord between thermodynamic models and experiments such as those illustrated in **Figure 7** reflects some more general features of these problems that transcend the specific model implemented to describe the transcription process (45).

2.4. The Regulatory Playground

The ideas set up in the specific case above for the regulatory architecture of simple repression can be used in all sorts of other settings. Indeed, as noted in the introduction, a broad range of circuits leading to switch-like and oscillatory behavior have been mathematicized and wired up in living cells (46–50). Several prominent and interesting examples that illustrate the framework spelled out above are shown in **Figure 10**. For the cases considered here, we now focus on the dynamics of the proteins rather than the mRNA molecules to give a flavor of the way that the same occupancy ideas have been used in other contexts. When the words and cartoons used to describe these

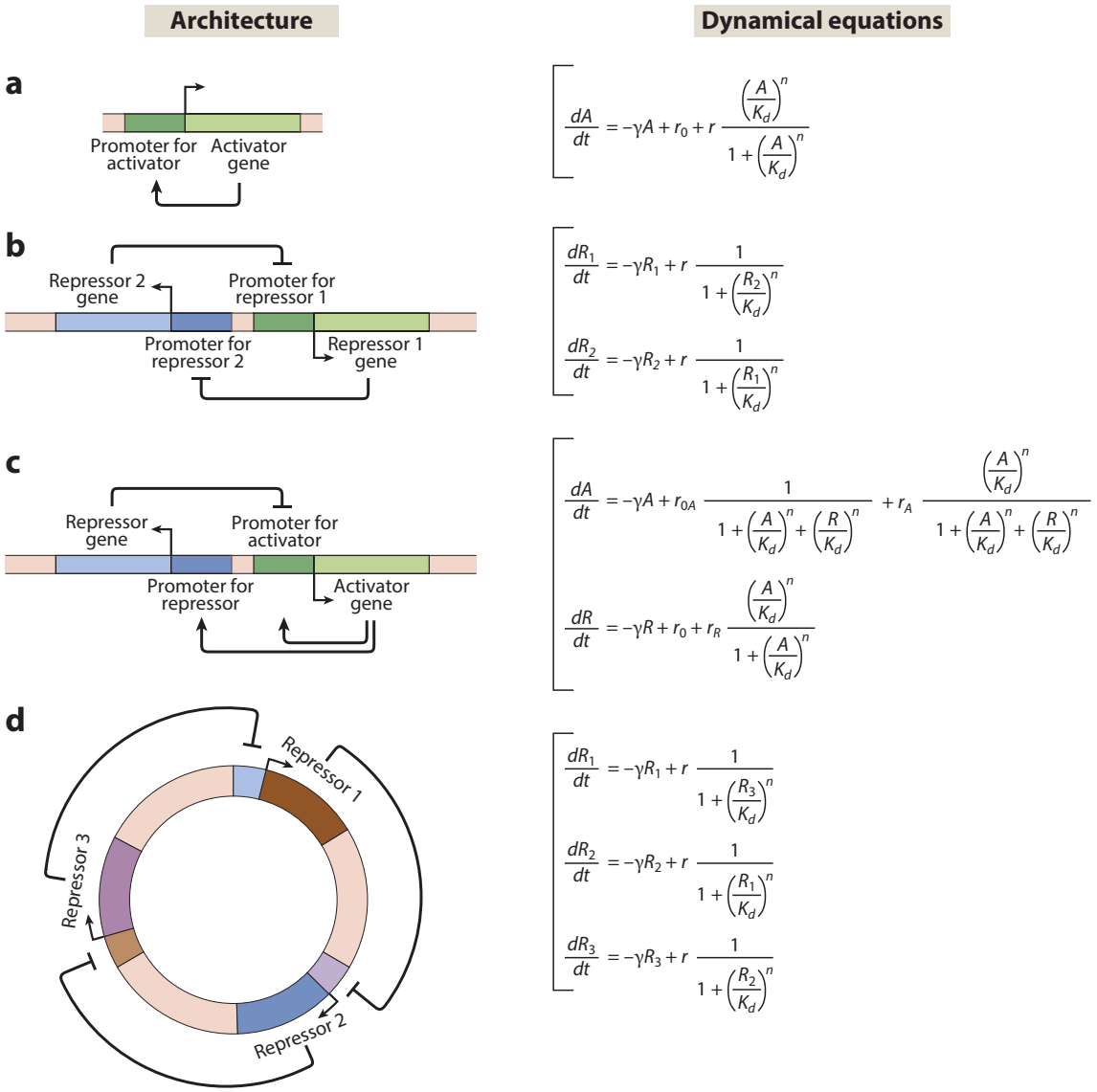


Figure 10

Regulatory architectures subjected to the thermodynamic framework. The first example is the self-activating switch. The second example is mutual repression, which leads to a genetic toggle switch. The third example is a two-component oscillator built from an activator and a repressor. The fourth example shows a genetic oscillator known as a repressilator. Detailed descriptions of all of these architectures can be found elsewhere (51).

circuits are mathematicized, the resulting structure for the governing dynamics nearly invariably takes the form

$$\frac{dA}{dt} = -\gamma A + r p_{bound}(A), \qquad 26.$$

where we consider the production rate of an activator molecule (A). The main point I am trying to emphasize is the use of occupancy functions that characterize the probability of the various states of the promoter.

Sometimes for simplicity when writing the probabilities that showed up abstractly in Equation 3, it is mathematically convenient to resort to a strictly phenomenological treatment of those probabilities by using a so-called Hill function (52). For example, if we consider the self-activating switch of **Figure 10*a***, we have

$$\frac{dA}{dt} = -\gamma A + r_0 + r\frac{\left(\frac{A}{K_d}\right)^n}{1 + \left(\frac{A}{K_d}\right)^n}, \qquad 27.$$

as shown in the first of the examples in **Figure 10**. The first term on the right captures the protein degradation process, the second term reflects basal production at rate r_0, and the final term reflects the feedback of the circuit because it has the property that the more A there is, the higher the rate of protein production. This circuit is a very nice exercise to explore how a simple dynamical equation reflecting elementary processes of protein dynamics can give rise to switch-like behavior in the sense that the steady-state value of the protein level can be either low or high (51).

Figure 10 shows a number of other examples, each of which can exhibit interesting dynamical behaviors depending on the specific choices in parameter space. For example, the second example in the figure shows a different version of a genetic switch (46, 49). **Figure 10*c*** shows an elementary two-component oscillator in which an activator activates both itself and a repressor. The repressor, in turn, represses the activator. For the right choice of parameters, this circuit will exhibit oscillations (51, 53, 54). The final example is a famous example of an oscillatory circuit (47). Again, our reason for highlighting these examples is to make it clear that the probabilistic language of occupancy suffuses the study of regulatory decision making in vivo; hence, an important part of trying to make the field quantitatively rigorous is to explore the limits and validity of the probabilistic description of occupancy using Boltzmann statistics or steady-state probabilities.

3. TWO-FACED MOLECULES: THE JANUS EFFECT

3.1. Regulating the Regulators

One of the most interesting things about living organisms is that they react to environmental cues. Bacteria can detect changes in their environment ranging from small concentration differences in some metal atoms or ions to changes in the osmotic pressure coming from the surrounding medium. Eukaryotic cells have similar capacities for detecting everything from the stiffness of the substrate they are growing on to the presence of signaling molecules in their environment. Indeed, our precious senses all rely on the ability of our cells to respond to external stimuli.

One of the mysteries resolved by modern biology is the determination of the kinds of molecular mechanisms that make it possible to detect and respond to all these different kinds of stimuli.

The general idea is now just five decades old and goes under the name of allostery, and refers to the fact that many biological macromolecules can switch back and forth between two conformations: an active and an inactive state (55–58). The relative probability of these two states is controlled, in turn, by the binding of some ligand to the allosteric molecule (59, 60). In this section of the article, we explore how these ideas can be converted into statistical mechanical language and then applied to the problem of cellular decision making, although the scope of the allostery concept is much broader than the limited application considered here in the context of transcriptional regulation (59, 60).

3.2. Biology's Greatest Model

Physics is defined by its many simple and elegant models. For mechanical systems near equilibrium, we invoke the harmonic oscillator model. The study of critical phenomena passes invariably through an analysis of the Ising model. For the specific heats of crystalline solids, we turn to the models of Einstein and Debye. The random walk model describes not only the intuitive case of diffusion but also the statistics of polymer conformations. And the list of simple and elegant overarching models goes on.

One of the most interesting and surprising features of these models is their broad reach. The harmonic oscillator tells us not only about springs but also, for example, about RLC circuits and the absorption of radiation by matter. Similarly, we use the Ising model in settings as diverse as the study of liquid-gas transitions, of the order-disorder transition in alloys, and of magnetism. Is it realistic to expect the same kind of broad reach for simple models in the biological setting?

Shortly after the elucidation of the allostery concept, a statistical mechanical model of allosteric molecules was advanced that elegantly captures several of the key conceptual features present in allosteric molecules, namely, the existence of several conformational states corresponding to the active and inactive states as well as a different affinity for binding some control ligand when in those different states. That model, sometimes referred to by the initials of its founders Monod, Wyman, and Changeux (hence, the MWC model), has now been applied to a host of different problems (57, 59–66). At roughly the same time, a second class of models was introduced that differs in subtle ways from the MWC model, which need not concern us here because we are trying to make a broader point about molecules that have several conformational states with different activity (67). Here, in Section 3, I illustrate how such models work from the statistical mechanical perspective and the kinds of things they can be used to understand.

3.3. Repression Revisited

Although the model described in the previous section tells us that a very useful control parameter is the number of transcription factors, often both in the lab and in the real world, cells respond not by changing the number of transcription factors but rather by changing the number of active transcription factors. How does this work? One of the interesting applications of the allostery concept in general and the MWC model in particular is to the study of transcriptional regulation. The key point is that both repressors and activators, the molecules considered in detail in the previous section of the paper, are often subject to environmental control by the binding of ligands to the transcription factor. For example, imagine a repressor molecule that controls the production of a protein that carries out an enzymatic reaction on some substrate. When that substrate is present in high quantities, the repressor is in its inactive (i.e., ligand bound) state. However, when the concentration of the ligand is low, it is no longer bound to the repressor molecules and hence the repressor is active and serves to turn down the level of expression. In other words, allostery

provides a coherent mechanism of biological feedback that can be implemented mathematically in the context of the MWC model.

How do we turn these ideas into a statistical mechanical framework for thinking about induction of genes by some downstream substrate? **Figure 11** shows one approach, which is to imagine that our collection of repressors can be separated into two populations, one of which is active and the other of which is induced and hence less active. This means that we need to generalize our original picture of simple repression to include two distinct classes of repressor, one corresponding to the active repressor (R_A) binding and the other to the reduced binding that accompanies binding of the inducer to the repressors (R_I). Clearly, we have a constraint on the total number of repressors of the form $R_A + R_I = R_{tot}$. Although in principle the model allows for the repressor to be in the active state while bound to a ligand and, similarly, to be in the inactive state, although not bound to ligand, we adopt the simplifying assumption (exemplified in **Figure 11**) that the active state of the repressor has no bound ligand and that the inactive state of the repressor occurs only when the ligand is bound.

Using the states and weights shown in **Figure 11**, we can write the probability that the gene is on as the ratio of the statistical weight of the state with polymerase bound to the sum over the statistical weights of all states, resulting in

$$p_{on} = \frac{\frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}{1 + \frac{R_A}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd,A}} + \frac{R_I}{N_{NS}} e^{-\beta \Delta \varepsilon_{rd,I}} + \frac{P}{N_{NS}} e^{-\beta \Delta \varepsilon_{pd}}}. \tag{28.}$$

If we introduce the simplifying notations $p = (P/N_{NS}) e^{-\beta \Delta \varepsilon_{pd}}$, $r_A = (R_A/N_{NS}) e^{-\beta \Delta \varepsilon_{rd,A}}$, and $r_I = (R_I/N_{NS}) e^{-\beta \Delta \varepsilon_{rd,I}}$, then we can rewrite our expression for the gene being in the on state as

$$p_{on} = \frac{p}{1 + r_A + r_I + p}. \tag{29.}$$

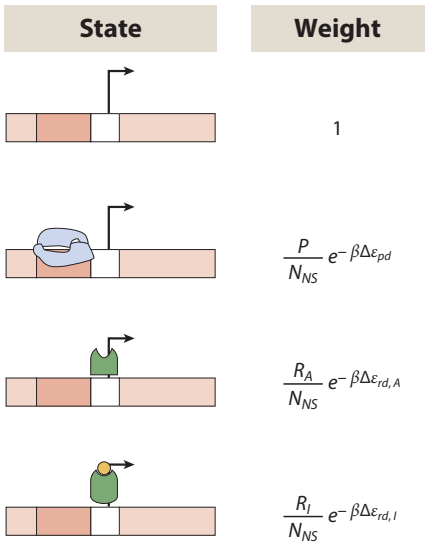


Figure 11

States and weights for a dual population of repressors. For the simple repression motif considered here, there are four states of the promoter: empty, occupied by RNA polymerase, occupied by uninduced repressor, and occupied by induced repressors.

As noted earlier, in many instances, it is most convenient for the purposes of comparison of experiments to reckon the fold-change (see Equation 11), which for this case is given by

$$\text{fold-change} = \frac{1}{1 + r_A + r_I}, \tag{30.}$$

where we have once again invoked the weak-promoter approximation ($p \ll 1$).

We can simplify our analysis by making the assumption that the inactive repressors have their K_ds increased so much by inducer binding (i.e., $\Delta\varepsilon_{rd,I} \ll \Delta\varepsilon_{rd,A}$) that they no longer bind the operator DNA at all (i.e., $r_I \ll 1$). In this case, the expression for the fold-change simplifies to

$$\text{fold-change} = \frac{1}{1 + r_A}. \tag{31.}$$

Of course, this is more easily interpreted if written in terms of the total number of repressors and the concentration of the inducer. To that end, we note that we can rewrite the expression as

$$\text{fold-change} = \frac{1}{1 + \frac{p_A(c)R}{K}}, \tag{32.}$$

where we have used the fact that $R_A = p_A(c)R$, with $p_A(c)$ itself defined as the probability of the repressor being in the active state as a function of the concentration c of the inducer. K is the effective (dimensionless) dissociation constant of active repressors to DNA. To make further progress, we must determine how $p_A(c)$ depends on the concentration of the inducer (c), and this is where the allostery models that are the subject of this part of the paper come in. Note that, as with many of the biological applications of statistical mechanics, we invoke these ideas with a separation of timescales in mind. That is, the binding and unbinding of the inducer from the repressor molecule and the subsequent switching between active and inactive states is much faster than the residence time of the repressor on the DNA once it is bound. Essentially, we imagine the equilibration of repressor and inducer to be instantaneous when compared with the timescale over which the repressor is bound or unbound on DNA.

We begin with the simplest model in which we imagine our repressor molecule to have one binding site for an inducer molecule. In this setting, this means that there are four states available to the repressor molecule, as shown in **Figure 12**. Specifically, there are two active states in which the molecule can repress and two inactive states in which it is assumed to no longer bind the operator DNA and hence cannot repress. The two states for each activity level correspond to the cases in which the inducer is either bound or not.

Given the states and weights shown in **Figure 12**, we can compute the activity of the repressor as a function of inducer concentration (c) in the form

$$p_A(c) = \frac{\left(1 + \frac{c}{K_d^A}\right)}{\left(1 + \frac{c}{K_d^A}\right) + e^{-\beta\varepsilon}\left(1 + \frac{c}{K_d^I}\right)}, \tag{33.}$$

where we introduce distinct dissociation constants K_d^A and K_d^I for the active and inactive states, respectively. We can now substitute this result back into Equation 32, which then yields an equation for the fold-change as a function of inducer concentration. Specifically, we have

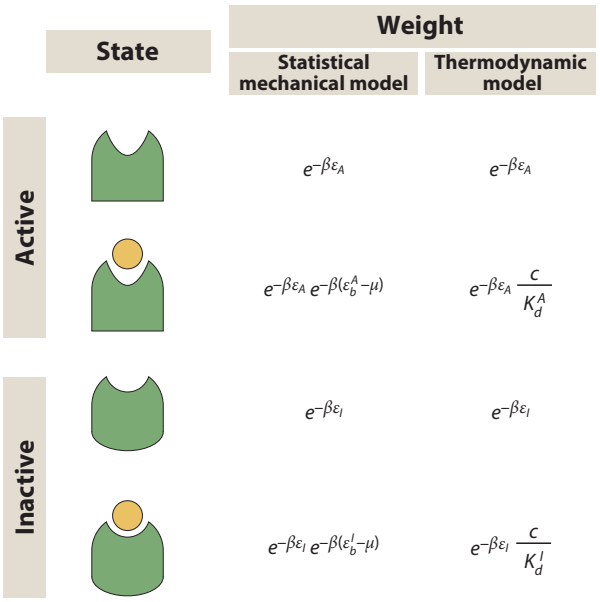


Figure 12

States and weights for the binding of inducer to repressor. The top two states correspond to active conformations of the repressor, and the bottom two states are the inactive conformation of the repressor. The statistical mechanical weight is written in terms of binding energies of ligand to the active (ε_b^A) and inactive (ε_b^I) states and the chemical potential μ, which characterizes the inducer, whereas the thermodynamic weights are written in terms of the K_ds of the active (K_d^A) and inactive (K_d^I) states and the inducer concentration c.

$$\text{fold-change} = \frac{1}{1 + \dfrac{\left(1+\frac{c}{K_d^A}\right)}{\left(1+\frac{c}{K_d^A}\right) + e^{-\beta\varepsilon}\left(1+\frac{c}{K_d^I}\right)}\frac{R}{K}}, \tag{34.}$$

where K_d^A and K_d^I are dissociation constants that characterize the binding of inducer to repressor, and K is an effective (i.e., dimensionless) dissociation constant that characterizes the binding of the active repressor to DNA.

We can rewrite the fold-change in a particularly interesting way that provides an opportunity for data collapse of broad collections of data from different mutants, as already shown in the settings of both chemotaxis and quorum sensing (62, 63). Specifically, we rewrite the fold-change in the form

$$\text{fold-change} = \frac{1}{1 + e^{-\beta F(c)}}, \tag{35.}$$

where $F(c)$ is defined as

$$F(c) = -k_B T\left(\log\frac{R}{K} + \log p_A(c)\right). \tag{36.}$$

This can be rewritten as

$$F(c) = -k_B T \left(\log \frac{R}{K} + log \frac{\left(1 + \frac{c}{K_d^A}\right)}{\left(1 + \frac{c}{K_d^A}\right) + e^{-\beta\varepsilon}\left(1 + \frac{c}{K_d^I}\right)} \right), \quad 37.$$

which is a parameter that I call the Bohr parameter because it is inspired by the families of binding curves seen in the binding of oxygen to hemoglobin. It is of interest to see to what extent our ideas will shed light on gene expression measurements.

Mutants have served as one of the most important windows to biological function. Further, different repressor mutants have been used to great effect for uncovering the allosteric properties of transcription factors (68). Recently, these strategies have been converted from a genetic tool into a biophysical tool. I like to think of this in much the same way that recent advances in condensed matter physics have been so tightly coupled to, for example, amazing advances in synthesizing new materials with all sorts of different magnetic and superconducting properties. And just as each new such material does not require an entirely new theory, neither does each of our mutants require some new theoretical description. Part of the distraction that is offered by a strictly structural view of allosteric molecules is that it focuses attention on the relative positions of different residues and ligands without pushing for a unifying perspective on all mutants at once. It is our hope that by virtue of the connection to the MWC framework, mutants can now be seen, for example, as arising either because of changes in free energy between the active and inactive states or through changes to the binding affinity of the inducer in the active and inactive states. Regardless of mutant type, the ambition is to exploit the same underlying mathematical description as a conceptual framework for explaining why the molecules behave the way they do and as a predictive framework for suggesting interesting new experiments.

Many transcription factors are more complicated than the monomeric repressors considered in **Figure 12**. For example, consider a repressor such as that shown in **Figure 13** that has two binding sites for inducers. In this case, our states and weights need to be expanded to account for the four distinct bound states corresponding to both the inactive and active states of the molecule. Given the states and weights, we can write the probability of the active state as

$$p_A(c) = \frac{\left(1 + \frac{c}{K_d^A}\right)^2}{\left(1 + \frac{c}{K_d^A}\right)^2 + e^{-\beta\varepsilon}\left(1 + \frac{c}{K_d^I}\right)^2}. \quad 38.$$

As before, we can also use this in the context of Equations 35 and 36 to write the fold-change itself as

$$\text{fold-change} = \frac{1}{1 + \frac{\left(1 + \frac{c}{K_d^A}\right)^2}{\left(1 + \frac{c}{K_d^A}\right)^2 + e^{-\beta\varepsilon}\left(1 + \frac{c}{K_d^I}\right)^2} \frac{R}{K}}, \quad 39.$$

where the formula is nearly identical to that provided in Equation 34, with the difference being that the terms involving the inducer concentration are squared, reflecting the fact that there are two binding sites for the inducer.

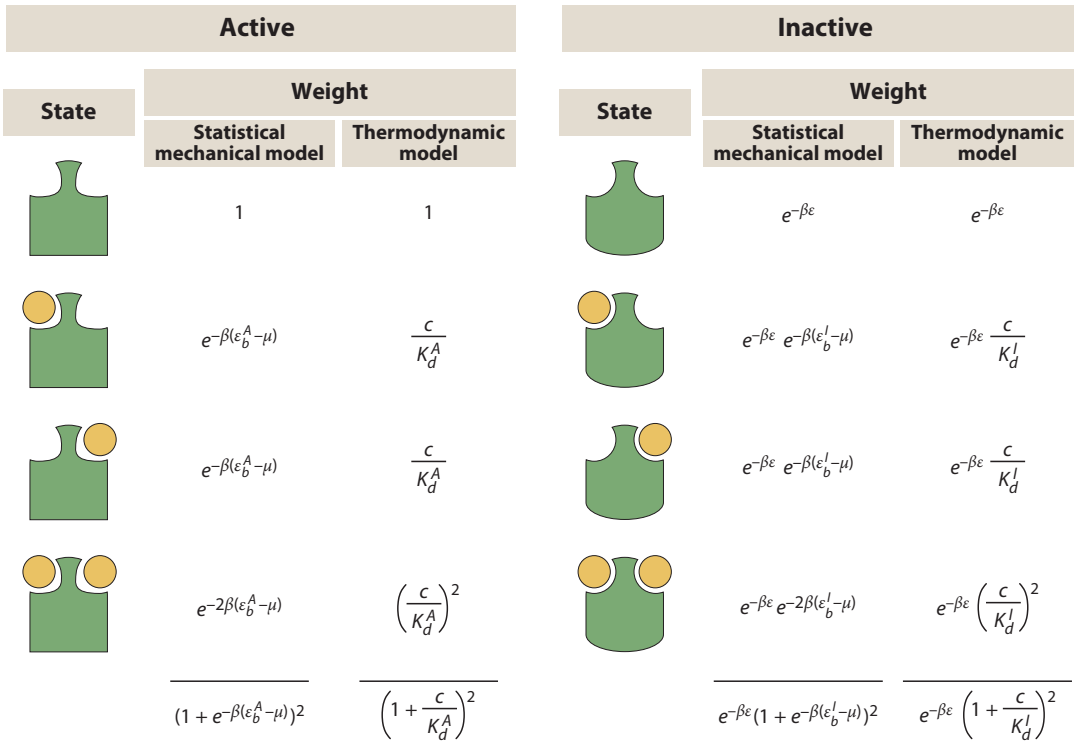


Figure 13

States and weights with an inducer that can bind at two sites on the repressor molecule. The notation is the same as that presented in **Figure 12**.

Although interesting as a statistical mechanical exercise, all of these abstractions are compelling only if they can tell us something about real experimental situations. The story of modern gene regulation owes much to the story of lactose metabolism in *E. coli*. In particular, in the heroic series of experiments in which the Lac repressor molecule was systematically mutated in order to determine which parts of the protein are responsible for DNA binding, for tetramerization, for inducer binding, and so on, the groundwork was laid for exploring repression as an allosteric phenomenon. Recently, this work has come full circle back to the systematic, quantitative in vivo study of regulation with a series of mutant repressor molecules (68, 69). The idea is to examine the response of a promoter controlled by the Lac repressor as a function of inducer concentration for a family of mutants of different kinds (68). Examples of the kind of data considered are shown in **Figure 14**. In this case, in addition to the wild-type repressor, the two mutants shown each have a single amino acid change in their DNA binding domain that alters the affinity of the repressor for DNA. From the standpoint of Equation 39, what this means is that all three data sets should be described by precisely the same parameters (i.e., ε, K_d^A, and K_d^I), with the exception of K, which characterizes the repressor-DNA interaction.

As noted above, the goal of the kinds of equilibrium models described in this article is to provide a unifying perspective on broad classes of biological phenomena in a way that suggests new experiments. We continue our analysis of the Lac repressor mutant data by examining the extent to which the data-collapse mentality suggested by Equation 37 can be brought to bear on this problem. As seen in **Figure 14*b***, the idea that all three data sets can be viewed from the point of view of the MWC framework is very appealing. Interestingly, altering the DNA binding strength causes the side effect of changes to the leakiness of the circuit as well as its dynamic range (59, 60).

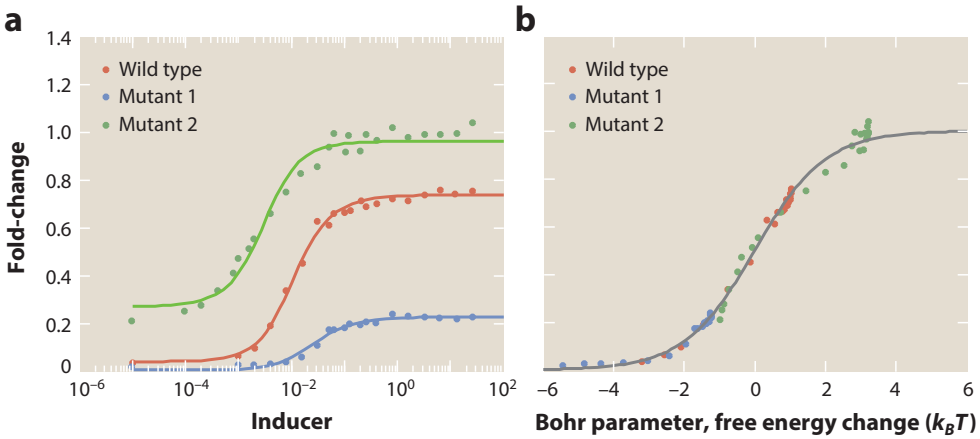


Figure 14

Gene expression from mutant repressors. (*a*) Experimental induction curves for wild-type Lac repressor and DNA-binding-domain mutants (68). Curves correspond to fits using Equation 39, where the only parameter that differs from one curve to the next is K, the effective dissociation constant for binding of active repressor to DNA. (*b*) Data collapse showing how the expression data for different repressor mutants collapse onto a single master curve by plotting expression data with respect to the Bohr parameter implied by Equation 39.

4. CONCLUDING PERSPECTIVE

Living organisms exhibit surprising phenomena at every turn, many of which nearly defy the imagination. That said, experimental advances in biology are unlocking secrets in a way that I would liken to the explosion of our understanding of the heavens after the telescope was invented. Despite the vast quantities of data generated about biological systems, the debate continues as to the extent to which our description of these systems can be quantitative and predictive, with an implicit implication that there might be things about living organisms that cannot be understood on the basis of physical law alone. The concept of this article has been to use several specific case studies illustrating how one particular approach to a class of molecular problems, namely statistical mechanics, provides guideposts for formulating a quantitative description of molecular actions and their resulting biological function. A second concept has been to begin to examine the limits and validity of these approaches and to ask what might be the kinds of new physics that will be needed to tackle these systems, which are characterized by a variety of challenges that make their statistical physics so hard: They have spatial and chemical heterogeneity, they are out of equilibrium, they exhibit sensitive dependence on single amino acid changes in the molecular

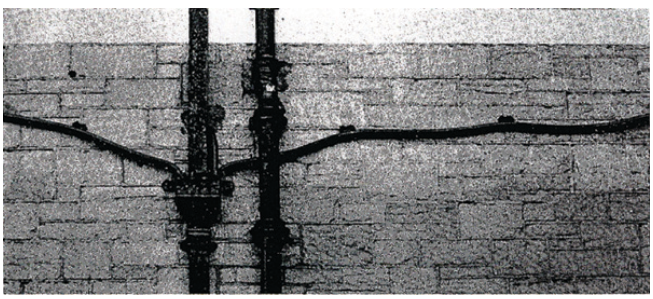


Figure 15

The sagging of a lead pipe on long timescales. Reprinted with permission from Reference 70.

substrates (i.e., enormity of space of mutants and their phenotypes), there is a strong degree of macromolecular crowding, etc.

At the deepest level, the knee-jerk reaction that rejects the use of statistical mechanics to describe cellular phenomena is right. Napoleon is in equilibrium. But these same kinds of worries apply in more mundane situations, such as in the use of Newton's First Law of Motion as a description of the equilibrium of a lead pipe in an old building (see **Figure 15**). Although such a pipe is clearly in mechanical equilibrium on short timescales, over longer timescales the process of creep results in permanent deformation as the pipe flows. Of course, the question of how to handle dynamics on the many different timescales that characterize biological systems is far trickier than in the context of mechanical creep of inanimate matter.

The use of statistical mechanical models like those discussed here is by no means limited to the analysis of transcription. A long list of examples from across biology have been subjected to the same kind of state and weight analysis given here, including the behavior of ion channels of all kinds, membrane receptors that mediate quorum sensing or bacterial chemotaxis, the distribution of proteins such as histones, polymerases, and transcription factors across genomes (as evidenced by a variety of recent sequencing methods), the accessibility of nucleosomal DNA, and beyond (64, 71, 72).

Even when not in equilibrium, many of these systems can be thought of as occupying distinct, identifiable states, and what they do functionally at a given instant depends upon which state they are in and on the rates of transitions between these states. Even in those cases, we are still faced with the same fundamental question from statistical physics: What are the probabilities? The simplest answer (null model) to that problem is that those probabilities are determined by Boltzmann and Gibbs distributions, but at best this represents a first approximation to the truth of how such systems behave. As shown here, the kinetic theory of transcription provides a second route to these probabilities, which, at least in some cases, are mathematically equivalent to those obtained from the statistical mechanical models. As future generations of researchers come to terms with the "test tube of the twenty-first century," statistical physics will enrich and be enriched by the study of living matter.

DISCLOSURE STATEMENT

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

ACKNOWLEDGMENTS

The chance to learn about this beautiful subject has served as the foundation of the happiest decade of my scientific life. This pleasure has in no small part derived from the privilege of learning from a broad array of distinguished scientists, including Lacra Bintu, James Boedicker, Rob Brewster, Robijn Bruinsma, Nick Buchler, Jean-Pierre Changeux, Robert Endres, Uli Gerland, Ido Golding, Terry Hwa, Tom Kuhlman, Jim Langer, Mitch Lewis, Sarah Marzen, Leonid Mirny, Alvaro Sanchez, Eran Segal, Franz Weinert, Ned Wingreen, and Jon Widom. This also includes the many students and postdocs that have joined in this adventure in my own laboratory. I need to particularly call out Hernan Garcia, Jane Kondev, Ron Milo, Nigel Orme, and Julie Theriot, with whom I have collaborated on decade-long book writing projects that have been among the most exciting experiences of my professional life. It is from these treasured collaborators that I have learned the most. The opinions and point of view expressed here are my own and those acknowledged above may not agree with a variety of things that they find in this article. It has also been a privilege to be entrusted

by the National Science Foundation, the National Institutes of Health, The California Institute of Technology, and La Fondation Pierre-Gilles de Gennes with the funds that make this kind of research possible, and for that I will be eternally grateful. Specifically, I am grateful to the NIH for support through award numbers DP1 OD000217 (Director's Pioneer Award) and R01 GM085286.

LITERATURE CITED

1. Hopfield JJ. 2013. *Annu. Rev. Condens. Matter Phys.* 5:1–13
2. Echols H, Gross C. 2001. *Operators and Promoters: The Story of Molecular Biology and Its Creators*. Berkeley: Univ. Calif. Press
3. Purnick PE, Weiss R. 2009. *Nat. Rev. Mol. Cell Biol.* 10:410–22
4. Mukherji S, van Oudenaarden A. 2009. *Nat. Rev. Genet.* 10:859–71
5. Bashor CJ, Horwitz AA, Peisajovich SG, Lim WA. 2010. *Annu. Rev. Biophys.* 39:515–37
6. Slusarczyk AL, Lin A, Weiss R. 2012. *Nat. Rev. Genet.* 13:406–20
7. Halder G, Callaerts P, Gehring WJ. 1995. *Science* 267:1788–92
8. Toettcher JE, Gong D, Lim WA, Weiner OD. 2011. *Nat. Methods* 8:837–39
9. Toettcher JE, Weiner OD, Lim WA. 2013. *Cell* 155:1422–34
10. Motta-Mena LB, Reade A, Mallory MJ, Glantz S, Weiner OD, et al. 2014. *Nat. Chem. Biol.* 10:196–202
11. Sourjik V, Berg HC. 2002. *Proc. Natl. Acad. Sci. USA* 99:123–27
12. Sourjik V, Berg HC. 2002. *Proc. Natl. Acad. Sci. USA* 99:12669–74
13. Hersen P, McClean MN, Mahadevan L, Ramanathan S. 2008. *Proc. Natl. Acad. Sci. USA* 105:7165–70
14. Mettetal JT, Muzzey D, Gomez-Uribe C, van Oudenaarden A. 2008. *Science* 319:482–84
15. Muzzey D, Gomez-Uribe CA, Mettetal JT, van Oudenaarden A. 2009. *Cell* 138:160–71
16. Phillips R, Kondev J, Theriot J, Garcia H. 2012. *Physical Biology of the Cell*. New York: Garland Press
17. Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B. 1994. *EMBO J.* 13:3348–55
18. Muller J, Oehler S, Muller-Hill B. 1996. *J. Mol. Biol.* 257:21–29
19. Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. *Cell* 123:1025–36
20. Kuhlman T, Zhang Z, Saier MH, Hwa T. 2007. *Proc. Natl. Acad. Sci. USA* 104:6043–48
21. So LH, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. 2011. *Nat. Genet.* 43:554–60
22. Chaikin PM, Lubensky TC. 2000. *Principles of Condensed Matter Physics*. Cambridge, UK: Cambridge Univ. Press
23. Phillips R. 2001. *Crystals, Defects and Microstructures*. Cambridge, UK: Cambridge Univ. Press.
24. Grigorova IL, Phleger NJ, Mutalik VK, Gross CA. 2006. *Proc. Natl. Acad. Sci. USA* 103:5332–37
25. Ko MS. 1991. *J. Theor. Biol.* 153:181–94
26. Peccoud J, Ycart B. 1995. *Theor. Popul. Biol.* 48:222–34
27. Record MTJ, Reznikoff W, Craig M, McQuade K, Schlax P. 1996. In Escherichia coli *and* Salmonella *Cellular and Molecular Biology*, ed. FC Neidhardt, R Curtis III, JL Ingraham, ECC Lin, KB Low, et al., pp. 792–821. Washington DC: ASM Press
28. Kepler TB, Elston TC. 2001. *Biophys. J.* 81:3116–36
29. Sanchez A, Kondev J. 2008. *Proc. Natl. Acad. Sci. USA* 105:5081–86
30. Michel D. 2010. *Prog. Biophys. Mol. Biol.* 102:16–37
31. Ackers GK, Johnson AD, Shea MA. 1982. *Proc. Natl. Acad. Sci. USA* 79:1129–33
32. Shea MA, Ackers GK. 1985. *J. Mol. Biol.* 181:211–30
33. Buchler NE, Gerland U, Hwa T. 2003. *Proc. Natl. Acad. Sci. USA* 100:5136–41
34. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, et al. 2005. *Curr. Opin. Genet. Dev.* 15:116–24
35. Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, et al. 2005. *Curr. Opin. Genet. Dev.* 15:125–35
36. Gertz J, Siggia ED, Cohen BA. 2009. *Nature* 457:215–18
37. Sherman MS, Cohen BA. 2012. *PLOS Comput. Biol.* 8:e1002407
38. Perrin J. 1990 (1923). *Atoms*. Woodbridge, CT: Ox Bow Press
39. Garcia HG, Phillips R. 2011. *Proc. Natl. Acad. Sci. USA* 108:12173–78
40. Brewster RC, Weinert FM, Garcia HG, Song D, Rydenfelt M, Phillips R. 2014. *Cell* 156:1312–23

41. Elf J, Li GW, Xie XS. 2007. *Science* 316:1191–94
42. Hammar P, Wallden M, Fange D, Persson F, Baltekin O, et al. 2014. *Nat. Genet.* 46:405–8
43. Shahrezaei V, Swain PS. 2008. *Proc. Natl. Acad. Sci. USA* 105:17256–61
44. Garcia HG, Sanchez A, Kuhlman T, Kondev J, Phillips R. 2010. *Trends Cell Biol.* 20:723–33
45. Frank SA. 2014. *J. Evol. Biol.* 27:1172–78
46. Gardner TS, Cantor CR, Collins JJ. 2000. *Nature* 403:339–42
47. Elowitz MB, Leibler S. 2000. *Nature* 403:335–38
48. Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. 2006. *Nature* 440:545–50
49. Laslo P, Spooner CJ, Warmflash A, Lancki DW, Lee HJ, et al. 2006. *Cell* 126:755–66
50. Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. 2007. *Science* 315:1716–19
51. Phillips R, Kondev J, Theriot J, Garcia HG. 2013. *Physical Biology of the Cell*. New York: Garland Sci. 2nd ed.
52. Frank SA. 2013. *Biol. Direct* 8:31
53. Guantes R, Poyatos JF. 2006. *PLOS Comput. Biol.* 2:e30
54. Tsai TY, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JJE. 2008. *Science* 321:126–29
55. Gerhart JC, Pardee AB. 1962. *J. Biol. Chem.* 237:891–96
56. Monod J, Changeux JP, Jacob F. 1963. *J. Mol. Biol.* 6:306–29
57. Monod J, Wyman J, Changeux JP. 1965. *J. Mol. Biol.* 12:88–118
58. Changeux J-P. 2012. *Annu. Rev. Biophys.* 41:103–33
59. Martins BM, Swain PS. 2011. *PLOS Comput. Biol.* 7:e1002261
60. Marzen S, Garcia HG, Phillips R. 2013. *J. Mol. Biol.* 425:1433–60
61. Mello BA, Tu Y. 2003. *Proc. Natl. Acad. Sci. USA* 100:8223–28
62. Keymer JE, Endres RG, Skoge M, Meir Y, Wingreen NS. 2006. *Proc. Natl. Acad. Sci. USA* 103:1786–91
63. Swem LR, Swem DL, Wingreen NS, Bassler BL. 2008. *Cell* 134:461–73
64. Mirny LA. 2010. *Proc. Natl. Acad. Sci. USA* 107:22534–39
65. Hilser VJ, Wrabl JO, Motlagh HN. 2012. *Annu. Rev. Biophys.* 41:585–609
66. Endres RG. 2013. *Physical Principles in Sensing and Signaling*. Oxford, UK: Oxford Univ. Press
67. Koshland J. 1966. *Biochemistry* 5:365–85
68. Daber R, Sochor MA, Lewis M. 2011. *J. Mol. Biol.* 409:76–87
69. Daber R, Lewis M. 2009. *Protein Eng. Des. Sel.* 22:673–83
70. Frost HJ, Ashby MF. 1982. *Deformation-Mechanism Maps*. Oxford, UK: Pergamon Press
71. Schiessel H. 2003. *J. Phys. Condens. Matter* 15:R699–774
72. Narula J, Igoshin OA. 2010. *IET Syst. Biol.* 4:393–408

Annual Review of Condensed Matter Physics

Volume 6, 2015

Contents

Errata

An online log of corrections to *Annual Review of Condensed Matter Physics* articles may be found at http://www.annualreviews.org/errata/conmatphys