Chapter 1

Biology by the Numbers

Rob Phillips

California Institute of Technology

Science was once called "natural philosophy" and had as its purview the scientific study of all of nature. Increasing specialization led to a splintering of natural philosophy into a number of separate disciplines and one of the outcomes of this trend was that physics emerged largely as the study of inanimate matter. This peculiar state of affairs imposed an unnatural barrier which largely prevented physicists from seeing the study of living matter as part of their core charter. Similarly, the style of analysis favored in the life sciences was often descriptive isolating much of the biological mainstream from quantitative descriptions as the rule rather than the exception. An exciting outcome of the biological revolution of the last fifty years is that the study of living matter is emerging as a true interdisciplinary science that will enrich traditional physics and biology alike. I examine some of the philosophical underpinnings of physical biology and then illuminate these ideas through several case studies that highlight the interplay between quantitative data and the models set forth to greet them. One of the interesting outcomes of an analysis from the physical biology perspective is that topics that seem very distant biologically are next door neighbors in physical biology.

1.1. Introduction

Science has always been driven by the invention of new ways of observing and measuring the world around us. When Galileo turned his "tube" to the vault of the heavens, he discovered the phases of Venus and the moons of Jupiter.[1] By connecting telescopes to spectrometers or detectors that can "see" at wavelengths other than the visible, our view of the universe has been transformed again and again. Over the last fifty years while we have garnered an unprecedented view of inanimate matter at the scales of both the very large and the very small, another revolution has taken place in

the study of living matter. Indeed, in just over fifty years, biologists have gone from uncertainty as to whether the molecules of the gene are protein or nucleic acid[2] to an era in which high school students can reprogram bacteria to do their own bidding. As a result of the technological advances that have come on the heels of fundamental biological understanding, it is becoming increasingly possible to study living matter in a quantitative way resulting in the science of what one might call "physical biology".

One of the key tenets of physical biology is that quantitative data demands quantitative models. More specifically, as evidenced by casual perusal of almost any scientific journal reporting on the life sciences, data from biological experimentation is routinely quantitative, presented in the form of graphs illustrating the kinds of functional dependence that are the lifeblood of the interplay between theory and experiment in the physical sciences. One of the central arguments of physical biology as I see it is that theory must be practiced at a level commensurate with the experimental state of the art. That is, if those performing experiments are going to go to all of the trouble of generating reproducible, quantitative data, then the field must demand that theory be practiced in a way that can respond to this data with more than words and cartoons.

To get an impression of the kind of data that can inspire a new kind of interplay between theory and experiment in biology, Fig. 1.1 shows recent experiments involving the physics of genome management. In these two cases (both of which will be fleshed out in more detail later in the essay), there is an interplay between the informational and physical properties of DNA as the central molecule of heredity.[3] In the first case, the physical properties of DNA as a charged, semiflexible polymer place physical constraints on the way that viruses can pack their genomes within the tight confines of a viral capsid.[4] As a result of these physical effects, the virus has to resort to the services of an ATP-driven molecular motor in order to fully pack its genome. In the second case, DNA loops are formed when a transcription factor binds simultaneously to two sites near a promoter.[5] The outcome of the measurement in this case is that the extent of gene expression depends sensitively upon the length of the DNA loop between the two sites.[6] The second of these experiments convincingly demonstrates that reproducible, quantitative data can be gotten from the *in vivo* setting as well as from their *in vitro* counterparts.

Both of these cases implicitly involve one of the key injunctions of physical biology, namely, the existence of some "tuning variable" that can be varied systematically both experimentally and in the models that are put

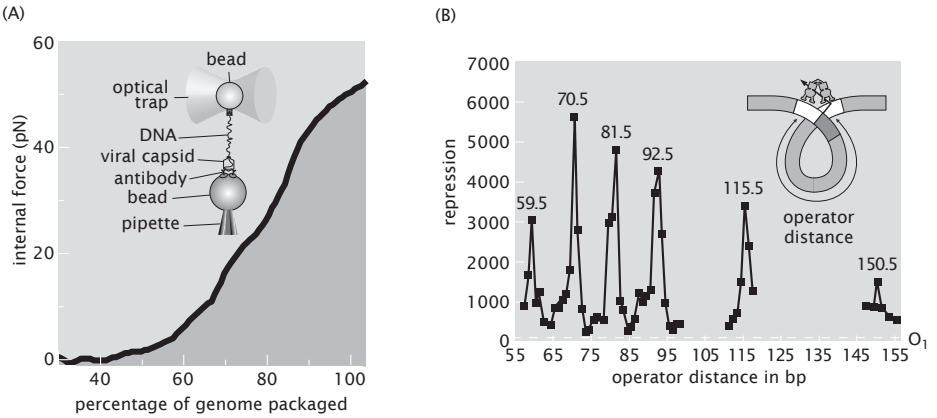

Fig. 1.1. Quantitative data demands quantitative models. Data from two experiments involving the *physical* properties of DNA. (A) Force build up during packing of a viral genome into the viral capsid. The inset shows a schematic of the optical tweezers experiment used to measure the force buildup during DNA packaging.[4] (B) Repression in the *lac* operon as a function of the distance between two Lac repressor binding sites. The inset shows a schematic of the binding of Lac repressor to DNA.[6]

forth to greet the data. As will be shown later in the paper, in the viral case, the genome length itself can be used as a tuning variable, whereas in the gene expression example, it is similarly the DNA length, but this time the distance between two binding sites. Generally, it is not the absolute magnitudes of the numbers that show up in these experiments (and the corresponding models) that are important. Rather, it is the ability to understand how these numbers vary as some key tuning variable is varied. For example, as shown in Fig. 1.1(B), by tuning the length of the DNA loop in one basepair increments, the relative strength of repression is varied considerably. Though the *absolute* value of repression may be out of reach of simple model building, the scaling with length provides a chance to test our understanding of the regulatory process.

Traditionally, much of biological understanding has been captured through subtle cartoons that represent a careful decision about which features of a problem are important and which are not. On the other hand, this kind of cartoon-level understanding is inadequate as a response to well characterized quantitative data like that described above. In the remainder of the essay, I will argue how we can go beyond pictorial and verbal

descriptions and use quantitative analysis as a test of understanding that is even more stringent. This kind of thinking has already been a limited but powerful part of the scientific study of life for more than a century with one of the most powerful examples being the analysis of propagation of action potentials with giants like Helmholtz measuring their velocity and Hodgkin and Huxley dissecting their mechanism.[7–10] One of the outcomes of this kind of analysis is that it can sharpen the kinds of questions that can be asked about a given problem and suggest new lines of experimental investigation that can't even be conceived in the absence of a quantitative framework.

1.2. Order of Magnitude Biology

One of the main obsessions that is passed along to students when they are first learning science is to check their units. In freshman chemistry, equations with long strings of unit conversions routinely navigate students between the number of grams of a particular reactant mixed in some solution and the energy released in the resulting reaction measured in kilojoules, for example. The reason teachers harp on students to "check their units" is that these unit checks constitute the first line of defense in the sanity check to make sure that the results make intuitive sense.

A less formulaic, but deeper class of sanity checks is offered by the arithmetic of order of magnitude estimates. The idea of such estimates is to find out if the magnitudes of the quantities in question are reasonable. I remember a student once computing the deflection of a bridge due to the weight of a train crossing the span and finding a result that was 10 to some large power (maybe 10^8) with units measured in meters! Clearly this result failed the order of magnitude sanity check, but the homework was still turned in. Aside from ferreting out outright errors, the more powerful use of simple estimates is their ability to tell us if we have the right factors in play.

There is a long tradition of the value of order of magnitude estimates in the service of both physics and biology. Indeed, one of the stories that epitomized the physicists of the mid-twentieth century concerns Enrico Fermi who asked his students in a graduate class to estimate the number of piano tuners in Chicago. However, this approach based upon getting numbers out at the end was a serious part of an approach to physics that has been argued as one of the core reasons for the strength of American physics after the second world war. The argument goes "In the United States, theory and

experiments were always housed under the same roof in one department. And perhaps more so than everywhere else, in the United States physics was about numbers, and theories were deemed to be algorithms for getting the numbers out."[11] More recently, Edward Purcell and Victor Weisskopf kept this tradition alive in the pages of the *American Journal of Physics* where they performed all sorts of interesting estimates in a set of papers that appeared over several years.[12–15]

This tradition continues today with one of the most useful classes in the physics curriculum at Caltech entitled "Order of Magnitude Physics". What is especially fun about this course is that the instructors (primarily Peter Goldreich and Sterl Phinney) have an open question policy in which they are willing to make an order of magnitude estimate of anything - the heights of mountains, the deflection of the Keck telescope mirrors under their own weight or the energy consumption of the human brain. My argument here is that this is definitely more than fun and games (though it definitely is fun) and serves as a self-conscious way to begin to understand what a given problem is about.[16] One of my favorite examples of the power and subtlety of order of magnitude estimates is the simple calculation that permits an order of magnitude determination of the lattice parameter, cohesive energy and bulk modulus of a metallic solid by treating it as an "electron gas".[17,18]

Perhaps the most famous example of this kind of *eureka* moment resulting from simple estimates was experienced by Newton when he realized that the ratio of the acceleration of a falling body near the Earth and the acceleration of the moon as it "falls" towards the Earth are inversely proportional to the square of their distances from the center of the Earth. It is probably not an exaggeration to say that this simple estimate was foundational in the discovery of the law of universal gravitation.[19] Newton himself noted "I deduced that the forces which keep the Planets in their Orbs must be reciprocally as the squares of their distances from centers about which they revolve: & thereby compared the force required to keep the Moon in her Orb with the force of gravity at the surface of the Earth, & found them answer pretty nearly."[19] To see the math unfold for ourselves, we need simply use the familiar formula $s = \frac{1}{2}gt^2$ and compare the distance fallen by an object at the surface of the Earth to the distance that the moon falls in the same time at its distance of nearly 60 Earth radii away from the center of the Earth. What we find is that in one second, an object right near the surface of the Earth will fall roughly 4.9 m, while the moon will "fall" 1.3×10^{-3} m. The ratio of these two distances is roughly 1/3700,

the "pretty nearly" to 1/3600 that Newton speaks of.[20] To see a more detailed and beautiful exposition of Newton's "Moon Test" see pgs. 357-361 of Chandrasekhar.[21]

It is exciting to hear Newton carry out this estimate in his own words.[22] Frankly, this estimate still just fills me with the thrill of seeing a deep and wonderful result revealed on the basis of a simple calculation and a few elementary facts. Proposition IV, Theorem IV of Newton's *Principia* says: "if we imagine the moon, deprived of all motion, to be let go, so as to descend towards the earth with the impulse of all that force by which it is retained in its orb, it will in the space of one minute of time, describe in its fall $15\frac{1}{12}$ *Paris* feet... Wherefore, since that force, in approaching to the earth, increases in the proportion of the inverse square of the distance, and, upon that account, at the surface of the earth, is $60 \cdot 60$ times greater than at the moon, a body in our regions, falling with that force, ought in the space of one minute of time, to describe $60 \times 60 \times 15\frac{1}{12}$ *Paris* feet; and, in the space of one second of time, to describe $15\frac{1}{12}$ of those feet. And with this very force we actually find that bodies here upon earth do really descend; for a pendulum oscillating seconds in the latitude of Paris will be 3 *Paris* feet, and 8 lines $\frac{1}{2}$ in length, as Mr. Huygens has observed. And the space which a heavy body describes by falling in one second of time is to half the length of this pendulum as the square of the ratio of the circumference of a circle to its diameter (as Mr. Huygens has also shown), and is therefore 15 *Paris* feet, 1 inch, 1 line $\frac{7}{9}$. And therefore the force by which the moon is retained in its orbit becomes, at the very surface of the earth, equal to the force of gravity which we observe in heavy bodies there. And therefore the force by which the moon is retained in its orbit is that very same force which we commonly call gravity."[22] That is an estimate!

A fundamental tenet of the physical biology approach, in my view, is the need to further strengthen this tradition. Barbara McClintock's interesting biography is entitled *A Feeling for the Organism* and builds the case that the best way to study biological systems is to really have a feeling for the organism.[23] From the perspective of physical biology, this feeling for the organism amounts to having a sense of the sizes of molecules, organelles, cells and organisms, an intuition for the rates of biological processes and a developed quantitative sense of the forces and energies relevant to biology.

One of the most interesting and simple ways to get an intuitive feel for the cell is to perform the cellular census.[24–26] This serves as a prime case study in order of magnitude biology. There are many different ways to come at such an estimate, all of which essentially converge on the same

qualitative picture: the cell is highly crowded.[27–29] One simple way to see this is to start with the fact that cytoplasm is characterized by a protein density of roughly 300 mg/mL.[30] Alternatively, we can build our order of magnitude estimates around the fact that the cell is roughly 30% by mass macromolecule with roughly half of that coming from proteins. If we assume a "typical" protein with a mass of 300 kDa, it is left as an exercise to the reader to demonstrate that this implies roughly three million proteins in an *E. coli* cell.[31] This estimate can be used in turn to calculate that the mean spacing between proteins is less than 10 nm. The key point here is that simply by knowing a few elementary facts it is possible to perform an order of magnitude analysis of a vast array of features of cells and organisms. The outcome of such thinking is shown in fig. 1.2.

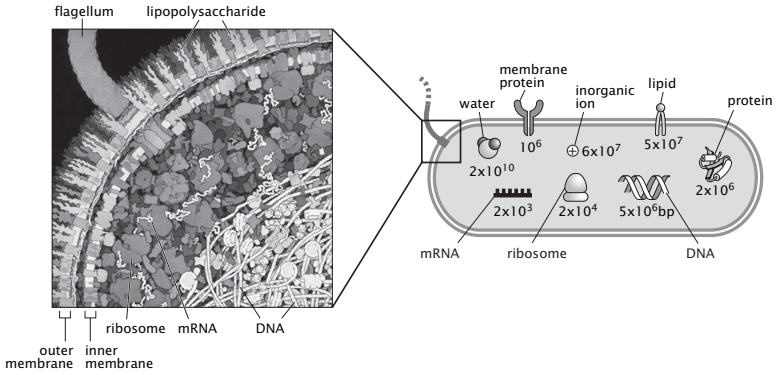

Fig. 1.2. Order of magnitude census of a bacterial cell. Goodsell's illustration of a bacterium shows the crowded nature of both the cytoplasm and the cell membrane.[27,32] The numbers on the right are *estimates* of the number of various classes of molecule in a bacterium.

We can complement the simple estimates on the crowded nature of the cytoplasm by reflecting similarly on the cell surface. Here too, there are many different ways to come to an estimate of the mean spacing between proteins on the cell surface. One of the simplest ideas is to use the fact that roughly 1/3 of genomes code for membrane proteins. In light of this rule of thumb, for a bacterium such as *E. coli*, this suggests that there are roughly 10^6 membrane proteins distributed roughly equally on the inner and outer membranes. A second way to come to the same basic conclusion is to appeal to recent measurements of the ratio of membrane protein mass to phospholipid mass. For example, in the *E. coli* membrane, for every 1 g

of phospholipid, there is roughly 2 g of protein.[33] This result can be used in turn to deduce the mean spacing between proteins. The upshot is that these kinds of order of magnitude estimates suggest a mean spacing between membrane proteins of less than 5 nm. One of the insights that emerges from such an estimate is that it is likely that there are strong physical interactions between membrane proteins which can alter their function.[34,35]

As noted above, these ideas are not new to biology. One of my favorite examples is offered in the beautiful Spiers Memorial Lecture of W. T. Astbury entitled "X-Ray Adventures Among the Proteins".[36] Astbury starts his paper in the kind of lively and personal style that has been systematically and sadly removed from almost all modern scientific prose with the words: "All scientific research is an intellectual adventure of course, and scientific thrills are among the best and highest kind of thrills. But nowhere do we feel this sense of adventure so much as when investigating living things and the complex bodies that take part in the life process. Far and away the most important and complex of these bodies are the proteins, and the problem of their structure and properties is, I think, the greatest scientific adventure of our times."[36]

Though we take the sequences, structures and functions of proteins and their macromolecular partners for granted, Astbury's 1938 lecture demonstrates that the way was not always clear. The part of the paper that especially intrigued me and is pertinent to the discussion of order of magnitude biology centers on his analysis of whether fibrous proteins (such as keratin) "are constructed to a common plan" with globular proteins. Astbury undertakes a simple estimate of the masses of the amino acid residues, noting "It is possible to calculate it from X-ray data and the density, however, without making use of the amino-acid proportions given by the chemistry".[36] The reader is encouraged to examine Astbury's paper to see how simple estimates were used as both a sanity check and a deductive tool as the attempt was being made to figure out what proteins are really like (again noting that just because something is now taught as a triviality in high school doesn't imply that it was always so obvious). The story of the long road to our understanding of proteins is described in Tanford and Reynolds' excellent book.[37]

An even more dramatic example of the way in which simple estimates can be biologically illuminating is to examine the biophysical underpinnings of fidelity. Biological polymerization during DNA replication, transcription and translation is very high with error rates lower than 10^{-4}.[38] From an order of magnitude perspective, one question that immediately comes to

mind is whether or not such high fidelity can be the result of thermodynamic specificity alone. For example, as was explored by both Hopfield[38] and Ninio,[39] if one computes the relative probability of binding the correct vs the incorrect tRNA during translation as a result of codon-anticodon pairing, this implies much higher error rates than are actually observed. The fact that the estimate does *not* jibe with measured error rates served as the basis of their analysis of kinetic mechanisms of discrimination.

The argument of this section is that, as Astbury (and many others) were inspired to do naturally, order of magnitude estimates can go a long way towards instructing us if we are on the right track in our thinking.

1.3. Case Studies in Physical Biology

After the first step of carrying out the relevant order of magnitude estimates, the next stage in one's analysis is often a more systematic examination of the problem from the point of view of both theory and experiment. In this section, I will describe three case studies that have been important to me personally that illustrate how such thinking might go. A different author could come up with a different set of examples (and the existence of many different examples is the whole point!) that would illustrate precisely the same point, namely, that the time is ripe to demand a rich and quantitative interplay between theory and experiment. Further, the essence of this new era of model building should explicitly aim for the kind of simplification and abstraction that intentionally attempts to ignore every detail of a given problem.

How Viruses Make New Viruses. Many of the most famous model systems from the history of molecular biology seem to have lost their luster in the biology community at large. However, in many cases, it is just these systems which have yielded so much biologically already that are optimally poised to serve as case studies in physical biology. Bacteriophage are among the most celebrated of model systems and were once referred to by Delbruck as the "atom of biology".[40] As already revealed in Fig. 1.1, bacterial viruses have served as a provocative model system for carrying out the kind of rich interplay between theory and experiment that is characteristic of physical biology. The measurements shown earlier in the paper considered the intriguing process of DNA packing. An interesting counterpart to the packing process is the ejection process whereby viral DNA is released from the bacteriophage once it has attached to the bacterium that is under siege.

We reasoned that by using genome length and salt conditions as tun-

ing parameters to dial in the pressure associated with the packed DNA, the rate of DNA ejection from these viruses would be altered. One of the most powerful tools in the physical biology arsenal are those coming from single-molecule biophysics. These experiments complement traditional bulk experiments by revealing the unique features of the trajectories of individual molecules. A beautiful single-molecule assay for viral ejection had already been developed[41] and we decided to tailor this assay to pose a well-formulated quantitative question: how does the rate of ejection depend upon genome length and on the salt conditions? As shown in Fig. 1.3, it is possible to watch individual phage particles eject their DNA one molecule at a time by fluorescently labeling the DNA as it emerges from the viral capsid. As noted above, the idea of the experiment is to examine how the rate of ejection depends upon tuning parameters such as the genome length. In the experiment shown here, two different genome lengths (the wild-type length of 48.5 kbp and a strain with a much shorter genome length of 37.5 kbp) are used to control the driving force for ejection.

Interestingly, in our own theoretical musings on this problem, we had a simple model of the ejection process based upon the idea that the friction experienced by the DNA is independent of how much DNA remains in the capsid.[43] Roughly at the same time our theoretical paper on the topic came out, my student Paul Grayson came to me with his experimental data showing that our model was overly simplistic. In particular, the friction experienced by the DNA as it is ejected from the capsid depends upon how much DNA remains in the capsid. However, from my point of view, the theory had done its job by suggesting these experiments in the first place. Now we are thinking hard about the nature of the friction associated with the packaged DNA and how it determines the rate of ejection. The very question itself is only meaningful when posed quantitatively.

How Cells Decide. One of the great episodes in the history of modern biology was the discovery of the idea of gene regulation. The simple idea is that there are certain genes whose job is to control the expression of other genes.[2,44] It is an interesting twist of history that the particular examples that led to the elucidation of the operon concept both involve DNA looping, a situation in which transcription factors bind at two sites on the DNA simultaneously and loop the intervening DNA. As shown in Fig. 1.1(B), beautiful, quantitative measurements reveal the subtle way in which transcriptional regulation depends upon physical factors such as the deformability of looped DNA.[6]

The study of transcriptional regulation has become impressively quanti-

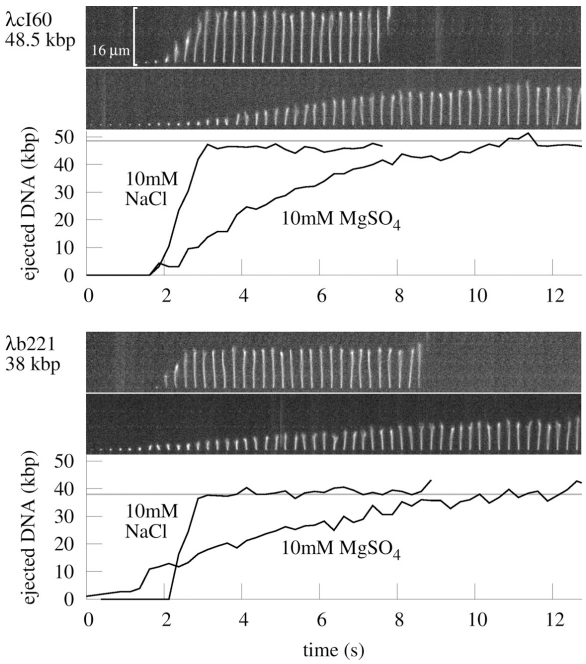

Fig. 1.3. Viral DNA ejection. Single-molecule measurements of the ejection process for two different choices of the viral genome length. The images show sequences of images of the same ejection event from one virion.[42]

tative. It is now possible to say how much expression occurs at a given time in the cell cycle and where within cells. As a result of these quantitative advances, it is incumbent upon those who are responding to this quantitative data to do so at a level that is similarly quantitative. One powerful tool for thinking about this class of problems is the "thermodynamic models" of gene regulation which make the simplifying but falsifiable assumption that the binding of RNA polymerase to the promoter of interest can be thought of as an equilibrium process.[45,46] Within this class of models, one computes the probability that the promoter is occupied and the predicted effect on gene expression is captured through a quantity known as the regulation factor.[47] This function gives the fold-change in gene expression as a function of key tuning variables such as repressor and activator concentrations, binding strengths of the transcription factor binding sites and quantities such as the free energy of looping. A wide range of regulation functions are shown in fig. 1.4 which shows how different regulatory architectures give

rise to different functional forms for the fold-change.

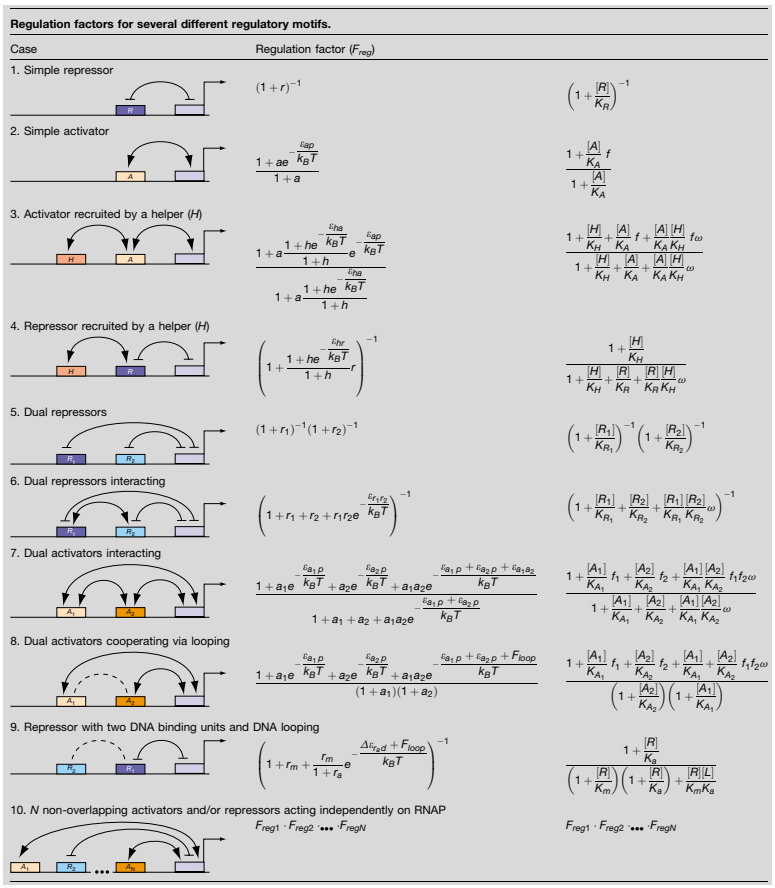

Fig. 1.4. Quantitative treatment of gene regulation. Each entry in the table corresponds to a different regulatory architecture. The mathematical expressions give the predicted "regulation function" for each motif and predict the fold-change in gene expression as a function of key parameters such as the concentration of transcription factors, the free energy of looping and the strengths of DNA-protein interactions.[47]

One of the interesting case studies that can be considered is that of DNA looping. As with the case of viral packing, attacking this problem from the physical biology perspective suggests that by tuning the DNA bendability (by controlling the sequence of the looped region), it is possible to probe transcriptional regulation in a way that explicitly appeals to model

predictions like those shown in Fig. 1.4. Experiments can be performed both *in vitro* (such as single-molecule looping experiments[48]) and *in vivo* single-cell measurements which directly test the predicted regulation factors and how they depend upon DNA flexibility, for example.

How Cells Detect Force. As another interesting example, this time from the cell biology and biophysics of membranes, we turn now to the subject of mechanosensation. Cells have a vast array of molecular components that couple to external forces.[49] One of the most palpable examples from everyday life is our sense of hearing which is based upon mechanosensors known as stereocilia which jiggle about in the hair cells of our inner ear as a result of sound waves impinging upon them. However, mechanosensation is ubiquitous and one of the most powerful model systems for exploring this phenomenon is offered by bacteria.

It has become possible to carry out a case study in physical biology on these problems because of the confluence of structural and functional information about the channels that mediate mechanosensation in the case of bacterial mechanosensitive channels. In this case, x-ray crystallography, crosslinking studies and other tools have resulted in an atomic-level picture of the structure in the closed state and hypotheses for what the channel looks like when open.[50–53] These results are complemented by functional measurements in which the current that passes through the channel is measured as a function of the load to which the surrounding membrane is subjected. Interestingly, by examining the gating properties of channels that are reconstituted in membranes with different lipid molecules, it has been found that the gating tension appears to depend sensitively on the properties of the surrounding membrane. Evidence for this effect is shown in Fig. 1.5. Though the details of how one might think about this kind of data using physical models are spelled out elsewhere,[54–57] for my purposes here the point of this data is to demonstrate yet again the way in which *scaling* of some observable property with some tunable parameter works out. In this case, the lipid tail length serves to alter the gating tension.

All three of the case studies highlighted here share the common feature that a particular parameter has been tuned that can be incorporated into our theory analysis and that elicit different biological function. By necessity, when biological questions are formulated in these precise quantitative terms, the questions are sharpened and the notion of what it means to understand a given phenomenon is tightened. Similarly, in each one of these case studies, it is possible to take the cartoon-level description of each problem and to recast it in mathematical terms.

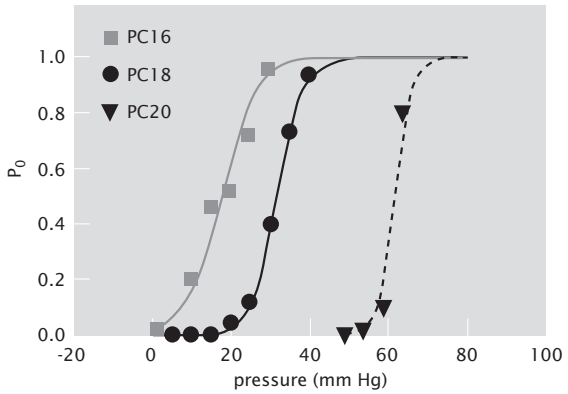

Fig. 1.5. Gating of mechanosensitive channels. The graph shows the probability the channel will be open (P_o) as a function of the suction pressure applied in the pipette. The critical pressure at which the mechanosensitive channel of large conductance opens depends upon the length of the lipid tail lengths in the surrounding membrane. The three curves correspond to lipid tail lengths of 16, 18 and 20.[58]

Mathematicizing the Cartoons. Like most ideas, the notion of translating visual representations into a corresponding mathematical formalism is not new. One of the greatest episodes in the history of physics was the realization of the unity of electricity, magnetism and optics. Much of the experimental and conceptual foundation for this revolution was ushered in by the experiments of Michael Faraday, captured in his *Experimental Researches in Electricity*.[59] As part of that revolution, Faraday introduced the concept of a "field" in a way that stuck and became one of the key features of the modern equipment of physics.

The experimental insights of Faraday concerning the relation between electricity and magnetism were described verbally and characterized conceptually by cartoons. It is into Faraday's world of lines of force permeating space that James Clerk Maxwell entered, equipped as he was with the mathematical tools of the theoretical physicist. By his own admission, Maxwell attached enormous importance to Faraday's experimental successes, commenting in the preface to his own classic *A Treatise on Electricity and Magnetism*, "before I began the study of electricity I resolved to read no mathematics on the subject till I had first read through Faraday's *Experimental Researches in Electricity*."[60] As a result of his reading of Faraday, Maxwell perceived that the key conceptual elements of the model of the electromagnetic field were already in place and that what was required

was to translate the cartoons and verbal descriptions of Faraday into the more familiar mathematical language of post Newtonian physics. Indeed, Maxwell himself characterized his successes thus: "I had translated what I considered to be Faraday's ideas into a mathematical form".[60]

Of course, analogies are almost never perfect, but I like to think that much of the beautiful and hard work that has gone into the biology of the last several hundred years is bringing us to the same sort of critical moment that was faced by Maxwell, and which was similarly illustrated by Kepler once Tycho Brahe had made his careful, quantitative measurements on the motion of Mars. In both of these cases, qualitative observations had given way to quantitative measurements. It was only in light of these measurements that it became possible to finally crack problems such as the motion of celestial bodies and the dynamics of electromagnetic fields. By way of contrast, these measurements raised the bar on what would constitute convincing theoretical understanding of these problems. The amazing explosion of understanding of all kinds about living organisms, much of which is captured diagrammatically in ways that show how different components are linked both informationally and physically serve as an invitation to the kind of translation into mathematical form described by Maxwell.

1.4. On the Virtues of Being Wrong

One of the challenges sure to be faced by those attempting to practice "physical biology" is a skepticism about the usefulness and correctness of theory and models. This skepticism comes in many different forms ranging from those who argue that theory offers nothing that the data can't already say for themselves, to those who are bothered by the omission of some feature of the problem, to those who dismiss models because they are "wrong" because they don't "agree" with some feature of the data. To take stock of the significance of these objections, we need to stop and examine what the goals of building simple models really are.

1.4.1. *Some "Wrong" Ideas We Teach in School*

I have recently heard it said that the goal of theory in biology is to be wrong. Though that might sound like an odd claim, I know just what was meant. One of the pillars of the physics curriculum is the study of thermodynamics and its microscopic partner science, statistical mechanics. Within the thermodynamic and statistical mechanical canon, one of the

most important topics is the study of specific heats, that is, the amount of heat that is required to raise the temperature of a substance by one degree. Interestingly, classical physics has a very concrete and universal prediction for the specific heat of crystalline solids that goes under the name of the law of Dulong and Petit which says that the specific heat is $3R$, where R is the universal gas constant.[61]

Interestingly, though this result is reasonably universal as a first cut at the high temperature specific heat of solids, the low temperature behavior is another matter altogether. Indeed, as the temperature drops, so too does the specific heat in direct contradiction with the law of Dulong and Petit. This anomaly was one of the great challenges to the ideas of classical physics and resulted in one of the many key contributions of Einstein to the development of quantum theory.[62] But here too, Einstein's model made drastic simplifying assumptions that resulted in inconsistencies with the scaling of the low-temperature specific heat with temperature. Should we dismiss Dulong-Petit and Einstein simply because they do not agree with the data over the entire range of measured values? As evidenced by a half-century or more of teaching in the physical sciences, the answer is clearly no. Similar statements can be made about the Lorentz model of optical absorption,[63] the Bohr model of the atom, the electron gas picture of metallic solids or the Ising model of magnetism.[64] However, in each of these cases the virtues of these models have far outweighed their vices and there is a marked lack of nuance in simply dismissing them as "wrong".

Not surprisingly, the utility of hypotheses that have later been shown to be incomplete (or even truly wrong, such as the phlogiston hypothesis which attributed a material reality to heat) have played an important role in biology as well. My argument is that a more generous interpretation of the words "right" and "wrong" would go a long way towards making the interplay between quantitative theory and experiment work better in the biological setting.

1.4.2. *Of Soups and Sparks*

A fascinating episode in the recent history of biology that makes the same point as the example about specific heats of solids described above is the nature of synaptic transmission. How do different neurons communicate with one another? Strongly held and distinct views on the nature of synaptic connections go back to Ramon y Cajal and Golgi and were played out in the mid twentieth century as the controversy of soups and sparks.[65]

In particular, the question being debated was whether or not communication between neurons was electrical (along the lines of the action potential) or rather involved chemical messengers. Both hypotheses had their proponents and were used as the basis for drawing conclusions about how various experiments would turn out. It seems like serious overreaching to me to dismiss this historical episode as one in which one group had it right and the other didn't, for such throwaway remarks ignore the value of suggestive hypotheses as part of the iterative search for a better picture of how nature works.

This view of the role of theory in biology was stated succinctly by Francis Crick who noted: "The job of theorists, especially in biology, is to suggest new experiments."[66] A model is necessarily a distillation and is used to illustrate something such as a way of thinking, or how a system depends upon some parameter or another. One of the highest compliments that can be paid a model is to say that it suggests further experimentation. Ultimately, even if the hypothesis turns out to be flawed or to lack sufficient generality, it should still be viewed as one of the engines that drives scientific progress.

The history of science seems to me to send the message that often when our preconceived notions are inconsistent with experiments, these are halcyon moments when we are maximally poised to learn something new. One of the teachers from whom I have learned the most is E. T. Jaynes, one of the pioneers who showed the connection between ideas from information theory and statistical mechanics. Jaynes was also a pleasure to be around not only because of his smarts, but because he seemed to me to live a maxim we have always passed along to our kids when out skiing - if you don't fall, you aren't trying hard enough or skiing on terrain that is steep enough. It is exciting to be associated with ideas that are no longer on the bunny slopes of academia, where the risk of falling is always present, but the probability of learning something is almost sure. Jaynes summarized his views on calculations being wrong thus: "if our calculation should indeed prove to be 'unsound,' the result would be far more valuable to physics than a 'successful' calculation!"[67] As noted above, the modern quantum theory owes its existence to at least two such "unsound" results, namely, the failure of the classical theory of the specific heats of solids and the so-called ultraviolet catastrophe which was a similar anomaly of classical physics in explaining blackbody radiation.

Interestingly, recent measurements on the packing of DNA in bacteriophage ϕ-29 have cast doubt on the ability of models like those described in

this essay to properly account for the observed forces.[68] In the spirit of this section, we look upon these recent challenges to the theory as precisely the kind of interplay between theory and experiment that helps move understanding forward. The immediate reaction is not one of trying to find some way to tweak the parameters in the models so that the data is "fit" once again. Rather, it is to look into the underlying assumptions of the model and to see which assumptions should be relaxed or reconsidered.

1.5. A Look Ahead

A skeptical response to the thrust of this essay can be framed from at least several different perspectives. First, if understanding living matter (i.e. cells and organisms and the molecules that make them tick) is the objective, will invoking the kind of rich interplay between quantitative theory and experiment advocated here teach us anything fundamentally new about the living world? Second, from the standpoint of the physical sciences, is the study of living matter going to teach us anything new about the physical laws that animate the world? Both of these are useful questions and we examine them in turn.

Will demanding a quantitative interplay between theory and experiment reveal anything new, deep or interesting about the living world? Ultimately, the only way to really find out is to try it and see. However, I would bet that the answer is yes. To see why, we need look no further than the examples highlighted in this essay. For example, it is not an exaggeration to say that the quantitative discrepancies between the classical description of specific heats and measured low-temperature values was one of the main seeds of the modern quantum theory. Similarly, attention to details in the features of atomic spectra led to the elucidation of key ideas in modern physics such as spin. In this sense, my thinking is guided by historical analogies like these that have repeatedly demonstrated that new insights are revealed by demanding a quantitative accounting of some phenomenon of interest. It is precisely the *failure* of the simple models that provided the experimental and theoretical impetus for the next generation of discoveries.

Equally interesting is the question of whether new insights into physics itself will be provided by the study of living matter. The themes of energy, information and geometry are some of the most constant threads running through the physical sciences, and yet, they serve as the cornerstone of many biological phenomena as well. There is something still mysterious and wonderful about the complexity of living organisms and it still feels

like a long way from our understanding of the "complexity" revealed by ordinary matter to the richness of the living world. Further, I can't shake a lingering disappointment in the impotence of conventional physics in the face of systems that are far from equilibrium and my own sense is that the best place for physics to look for clues about how to think about nonequilibrium is in the setting of living matter. The idea that other sciences have something to gain from biology was stated far more eloquently as a query in the title of a paper by Joel Cohen: "Is mathematics biology's next microscope, only better? Is biology mathematics' next physics only better?"[69] What Cohen is noting is the historic synergy between mathematics and physics where ideas from one would enrich the other and then back again. In particular, for several centuries, applications in physics were one of the main driving forces in the development of mathematics and advances in mathematics permitted the expansion of physics to new domains. Cohen is betting on a similar synergy with biology at the center this time.

Ultimately, one of the main arguments of this essay is that the hard work and deep insights of biologists on a variety of model systems makes this an opportune time to reexamine these systems with an eye to the type of interplay between theory and experiment more familiar from physics. By demanding a more stringent (i.e. quantitative) definition of what it means to understand a system, as with many previous episodes (such as the specific heat example given earlier in the essay), we are sure to learn new things about the workings of nature. Further, by carrying out this program, physical biology will strengthen the roots of both physics and biology themselves.

Educating the Next Generation of Life Scientists and Engineers. Another way in which physical biology must surely touch the future of both physics and biology concerns the way we educate students. The traditional physics curriculum is based upon a canon of topics that includes mechanics, electricity and magnetism, statistical mechanics and quantum mechanics. Will the applications of the ideas of physics to living matter become a part of this canon, or will physics remain stuck with the bizarre adiabatic wall that separates the study of inanimate from living matter? Similarly, will tomorrow's biology students come to view their training in physics and mathematics as a reasonable price of entry into the study of biology? In a report released by the National Academy of Sciences entitled "Bio2010", it was argued that the future of education in the life sciences must include a nod to the more quantitative aspects of biology.[70]

Applied Biology. One of the most amazing features of science is in

the form of unanticipated technological spin offs. For every fundamental science, ultimately, there seems to always be an engineering partner. Fundamental understanding of the quantum mechanics of confined electrons gave rise to parts of a vast array of engineering disciplines (materials science, electrical engineering, computer science, etc.). Ironically, one of the poster problems learned by every beginning quantum mechanic, namely, the particle in a box, has become a technological reality. Similarly, fundamental studies of the mechanics of gases, liquids and solids literally serve as the scientific underpinnings used in engineering the new Airbus A380 or the Boeing "Dreamliner". Mechanical engineers use these foundational ideas to examine the structural mechanics of wings. Anyone who has ever sat in the window seat of a large airplane such as a 777 or a 747 will know that the wing tips can suffer deflections in excess of a meter and it is our understanding of the elasticity of structures that ensures that these wing tip deflections are nothing more than uncomfortable. Similarly, engineering fluid mechanics provides insights into the generation of wing tip vortices and drag that can be used to increase fuel efficiency. Materials scientists use fundamental understanding of the thermodynamics of alloys to construct enormous single-crystal turbine blades to prevent the kind of fracture at material interfaces that were the cause of air disasters. In each of these cases, enlightened empiricism has been superseded by rational engineering based upon rigorous scientific underpinnings. We should expect a similar transformation of the engineering outgrowths of biology.

The mindset represented by "physical biology" is one in which the construction of the fundamental biological infrastructure to fuel an engineering discipline is front and center. The February, 1989 issue of "Physics Today" was a special feature on the many contributions of Richard Feynman to modern physics. Several features of this issue caught my eye. One was that Feynman considered his teaching to be his paramount achievement (but that is another story). A second intriguing feature of this issue was a photograph of the blackboard of Feynman at the time of his death which had various interesting notes that he had written to himself. Of the many interesting things written on his blackboard, the one that caught my attention perhaps more than all the others was "What I cannot create I do not understand". Physical biology begins to move in the direction of that definition of understanding by demanding that quantitative data is met with predictive, quantitative models. More to the point, the emergence of synthetic biology in which new networks, cells and organisms are constructed from scratch demonstrates that the study of living matter is well on its way

to measuring up to Feynman's definition of understanding.

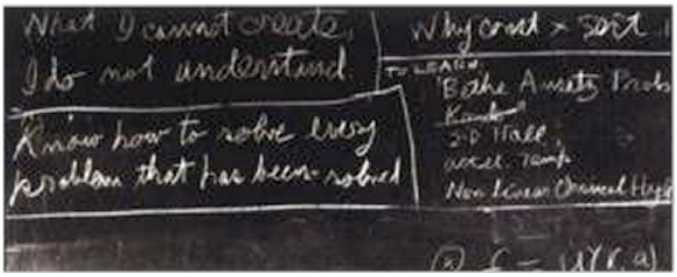

Fig. 1.6. Blackboard of Richard Feynman. Besides the quote mentioned in the text, the board also says "Know how to solve every problem that has been solved" and provides a list of things "to learn".

Acknowledgements

Many of the sentiments and ideas expressed in this essay have arisen as a result of my conversations with Jane' Kondev, Julie Theriot, Nigel Orme and Hernan Garcia in our work on our book *Physical Biology of the Cell*. I am grateful to each of them for the pleasure of working together and for having taught me so much. I should note, however, that many of the views expressed here did not pass muster with my friends and they should not be held responsible for any of my opinions that are found objectionable. Steve Quake and Jon Widom have both patiently instructed me on the interface between physics and biology. My referencing throughout the essay has not been exhaustive and is intended to provide an entry into the literature for the interested reader. I am grateful to the many others who have assisted me in my adventures in quantitative journalism as I have gone from one lab to another to learn about many of the amazing discoveries being made in the study of living matter. It is also a great pleasure to acknowledge the support of the NIH Director's Pioneer Award.

References

1. G. Galilei and S. Drake, *Discoveries and opinions of Galileo : including The starry messenger (1610), Letter to the Grand Duchess Christina (1615), and excerpts from Letters on sunspots (1613), The assayer (1623)*. (Anchor Books, New York, 1990).

2. H. F. Judson, *The Eighth Day of Creation*. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor ; New York, 1996).
3. H. G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P. C. Nelson, R. Phillips, J. Widom, and P. A. Wiggins, Biological consequences of tightly bent DNA: The other life of a macromolecular celebrity, *Biopolymers*. **85**(2), 115–30, (2007).
4. D. E. Smith, S. J. Tans, S. B. Smith, S. Grimes, D. L. Anderson, and C. Bustamante, The bacteriophage straight phi29 portal motor can package DNA against a large internal force, *Nature*. **413**(6857), 748–52, (2001).
5. B. Müller-Hill, *The lac Operon : a short history of a genetic paradigm*. (Walter de Gruyter, Berlin ; New York, 1996).
6. J. Müller, S. Oehler, and B. Muller-Hill, Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator, *J Mol Biol*. **257**(1), 21–9, (1996).
7. A. L. Hodgkin and A. F. Huxley, Resting and action potentials in single nerve fibres, *J Physiol*. **104**(2), 176–95, (1945).
8. A. L. Hodgkin, The ionic basis of electrical activity in nerve and muscle, *Biological Reviews of the Cambridge Philosophical Society*. **26**(4), 339–409, (1951).
9. A. L. Hodgkin and A. F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, *J Physiol*. **117**(4), 500–44, (1952).
10. A. L. Hodgkin and A. F. Huxley, Propagation of electrical signals along giant nerve fibers, *Proc R Soc Lond B Biol Sci*. **140**(899), 177–83, (1952).
11. S. S. Schweber, *QED and the men who made it : Dyson, Feynman, Schwinger, and Tomonaga*. Princeton series in physics, (Princeton University Press, Princeton, N.J.).
12. E. M. Purcell, Life at low reynolds number, *Am. J. Phys.* **45**, 3, (1977).
13. E. M. Purcell, The back of the envelope, *Am. J. Phys.* **51**(1), 11, (1983).
14. H. J. Bernstein and V. F. Weisskopf, About liquids, *Am. J. Phys.* **55**(11), 974–983, (1987).
15. M. Nauenberg and V. F. Weisskopf, Why does the sun shine?, *Am. J. Phys.* **46**(1), 23–31, (1978).
16. J. Harte, *Consider a spherical cow : a course in environmental problem solving*. (University Science Books, Mill Valley, Calif.).
17. A. L. Fetter and J. D. Walecka, *Quantum theory of many-particle systems*. International series in pure and applied physics., (McGraw-Hill, San Francisco, 1971).
18. D. Pines, *Elementary excitations in solids: lectures on phonons, electrons, and plasmons*. (W. A. Benjamin, New York, 1964).
19. R. S. Westfall, *Never at rest : a biography of Isaac Newton*. (Cambridge University Press, Cambridge [Eng]. ; New York, 1980).
20. A. P. French, *Newtonian mechanics*. (W. W. Norton, New York,, 1971).
21. S. Chandrasekhar, *Newton's Principia for the common reader*. (Clarendon Press ; Oxford University Press, Oxford [England] New York, 1995).
22. I. Newton, F. Cajori, R. T. Crawford, and A. Motte, *Sir Isaac Newton's*

Mathematical principles of natural philosophy and his System of the world. (University of California press, Berkeley, Calif.,, 1934).
23. E. F. Keller, *A feeling for the organism : the life and work of Barbara McClintock*. (W.H. Freeman, San Francisco, 1983).
24. S. Pedersen, P. L. Bloch, S. Reeh, and F. C. Neidhardt, Patterns of protein synthesis in *E. coli*: a catalog of the amount of 140 individual proteins at different growth rates, *Cell*. **14**(1), 179–90, (1978).
25. S. B. Zimmerman and S. O. Trach, Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*, *J Mol Biol*. **222**(3), 599–620, (1991).
26. J. Q. Wu and T. D. Pollard, Counting cytokinesis proteins globally and locally in fission yeast, *Science*. **310**(5746), 310–4, (2005).
27. D. S. Goodsell, Inside a living cell, *Trends Biochem Sci*. **16**(6), 203–6, (1991).
28. A. P. Minton, The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media, *J Biol Chem*. **276**(14), 10577–80, (2001).
29. K. Luby-Phelps, Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area, *Int Rev Cytol*. **192**, 189–221, (2000).
30. S. B. Zimmerman and A. P. Minton, Macromolecular crowding: biochemical, biophysical, and physiological consequences, *Annu Rev Biophys Biomol Struct*. **22**, 27–65, (1993).
31. F. C. Neidhardt, J. L. Ingraham, and M. Schaechter, *Physiology of the bacterial cell : a molecular approach*. (Sinauer Associates, Sunderland, Mass., 1990).
32. D. S. Goodsell, *Our molecular nature / the body's motors, machines, and messages*. (Copernicus, New York, 1996).
33. K. Mitra, I. Ubarretxena-Belandia, T. Taguchi, G. Warren, and D. M. Engelman, Modulation of the bilayer thickness of exocytic pathway membranes by membrane proteins rather than cholesterol, *Proc Natl Acad Sci U S A*. **101**, 4083–8, (2004).
34. H. Sprong, P. van der Sluijs, and G. van Meer, How proteins move lipids and lipids move proteins, *Nat Rev Mol Cell Biol*. **2**(7), 504–13, (2001).
35. T. Ursell, K. C. Huang, E. Peterson, and R. Phillips, Cooperative gating and spatial organization of membrane proteins through elastic interactions, *PLoS Comput Biol*. **3**(5), e81, (2007).
36. W. T. Astbury, X-ray adventures among the proteins, *Trans. Faraday Soc.* **34**, 378–388, (1938).
37. C. Tanford and J. A. Reynolds, *Nature's robots : a history of proteins*. (Oxford University Press, Oxford ; New York, 2001).
38. J. J. Hopfield, Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proc Natl Acad Sci U S A*. **71**(10), 4135–9, (1974).
39. J. Ninio, Kinetic amplification of enzyme discrimination, *Biochimie*. **57**(5), 587–95, (1975).
40. E. P. Fischer and C. Lipson, *Thinking about science : Max Delbrück and the*

origins of molecular biology. (Norton, New York, 1988), 1st edition.
41. S. Mangenot, M. Hochrein, J. Radler, and L. Letellier, Real-time imaging of DNA ejection from single phage particles, *Curr Biol.* **15**(5), 430–5, (2005).
42. P. Grayson, L. Han, T. Winther, and R. Phillips, Real-time observations of single bacteriophage lambda DNA ejections *in vitro*, *Proc Natl Acad Sci U S A.* **104**(37), 14652–7, (2007).
43. M. M. Inamdar, W. M. Gelbart, and R. Phillips, Dynamics of DNA ejection from bacteriophage, *Biophys J.* **91**(2), 411–20, (2006).
44. H. Echols and C. Gross, *Operators and promoters : the story of molecular biology and its creators.* (University of California Press, Berkeley, 2001).
45. G. K. Ackers, A. D. Johnson, and M. A. Shea, Quantitative model for gene regulation by lambda phage repressor, *Proc Natl Acad Sci U S A.* **79**(4), 1129–33, (1982).
46. N. E. Buchler, U. Gerland, and T. Hwa, On schemes of combinatorial transcription logic, *Proc Natl Acad Sci U S A.* **100**(9), 5136–41, (2003).
47. L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips, Transcriptional regulation by the numbers: models, *Curr Opin Genet Dev.* **15**(2), 116–24, (2005).
48. L. Finzi and J. Gelles, Measurement of lactose repressor-mediated loop formation and breakdown in single DNA molecules, *Science.* **267**(5196), 378–80, (1995).
49. P. G. Gillespie and R. G. Walker, Molecular basis of mechanosensory transduction, *Nature.* **413**(6852), 194–202, (2001).
50. E. Perozo, D. M. Cortes, P. Sompornpisut, A. Kloda, and B. Martinac, Open channel structure of mscl and the gating mechanism of mechanosensitive channels, *Nature.* **418**(6901), 942–8, (2002).
51. S. Sukharev, M. Betanzos, C. S. Chiang, and H. R. Guy, The gating mechanism of the large mechanosensitive channel mscl, *Nature.* **409**(6821), 720–4, (2001).
52. E. Perozo and D. C. Rees, Structure and mechanism in prokaryotic mechanosensitive channels, *Curr Opin Struct Biol.* **13**(4), 432–42, (2003).
53. S. Sukharev and A. Anishkin, Mechanosensitive channels: what can we learn from 'simple' model systems?, *Trends Neurosci.* **27**(6), 345–51, (2004).
54. C. Nielsen, M. Goulian, and O. S. Andersen, Energetics of inclusion-induced bilayer deformations, *Biophys J.* **74**(4), 1966–83, (1998).
55. P. Wiggins and R. Phillips, Analytic models for mechanotransduction: gating a mechanosensitive channel, *Proc Natl Acad Sci U S A.* **101**(12), 4071–6, (2004).
56. P. Wiggins and R. Phillips, Membrane-protein interactions in mechanosensitive channels, *Biophys J.* **88**(2), 880–902, (2005).
57. V. S. Markin and F. Sachs, Thermodynamics of mechanosensitivity, *Phys Biol.* **1**(1-2), 110–24, (2004).
58. E. Perozo, A. Kloda, D. Cortes, and B. Martinac, Physical principles underlying the transduction of bilayer deformation forces during mechanosensitive channel gating, *Nat Struct Biol.* **9**, 696–703, (2002).
59. M. Faraday, *Experimental researches in electricity.* (Dover Publications, New

York,, 1965).
60. J. C. Maxwell and J. J. Thomson, *A treatise on electricity and magnetism*. (Dover Publcations, [New York]), unabridged 3rd edition.
61. S. G. Brush, *Statistical physics and the atomic theory of matter : from Boyle and Newton to Landau and Onsager*. Princeton series in physics, (Princeton University Press, Princeton, N.J., 1983).
62. A. Pais, *"Subtle is the Lord– " : the science and the life of Albert Einstein*. (Oxford University Press, Oxford [Oxfordshire] ; New York, 1982).
63. F. Wooten, *Optical properties of solids*. (Academic Press, New York,, 1972).
64. N. W. Ashcroft and N. D. Mermin, *Solid state physics*. (Saunders College, Philadelphia, 1976).
65. E. S. Valenstein, *The war of the soups and the sparks : the discovery of neurotransmitters and the dispute over how nerves communicate*. (Columbia University Press, New York, 2005).
66. F. Crick, *What mad pursuit : a personal view of scientific discovery*. Alfred P. Sloan Foundation series, (Basic Books, New York, 1988).
67. E. T. Jaynes and R. D. Rosenkrantz, *E.T. Jaynes : papers on probability, statistics and statistical physics*. Synthese library ;, (D. Reidel ; Sold and distributed in the U.S.A. and Canada by Kluwer Boston, Dordrecht ; Boston Hingham, MA, 1983).
68. J. P. Rickgauer, D. N. Fuller, S. Grimes, P. J. Jardine, D. L. Anderson, and D. E. Smith, Portal motor velocity and internal force resisting viral DNA packaging in bacteriophage ϕ29, *Biophys. J.* (2008).
69. J. E. Cohen, Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better, *PLoS Biol.* **2**(12), e439, (2004).
70. N. R. C. U. C. on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century. Bio 2010 : transforming undergraduate education for future research biologists, (2003).