CellPress

Special Issue: Quantitative Cell Biology

Opinion

Theory in Biology: Figure 1 or Figure 7?

Rob Phillips[1,*]

The pace of modern science is staggering. The quantities of data now flowing from DNA sequencers, fluorescence and electron microscopes, mass spectrometers, and other mind-blowing instruments leave us faced with information overload. This explosion in data has brought on its heels a concomitant need for efforts at the kinds of synthesis and unification we see in theoretical physics. Often in cell biology, when theoretical modeling takes place, it is as a figure 7 reflection on experiments that have already been done, with data fitting providing a metric of success. Figure 1 theory, by way of contrast, is about living dangerously by turning our thinking into formal mathematical predictions and confronting that math with experiments that have not yet been done.

Trends

The rapid pace of experimental advance and acquisition of exciting news kinds of data in cell biology makes it ever more important to develop conceptual frameworks that unify and explain that data.

Mathematical theory forces us to formally state our thoughts in the same way that writing a computer program demands a precise statement of the underlying algorithm.

Theoretical models complement biochemistry, genetics, bioinformatics, and other frameworks for querying biological systems.

Theory allows us to sharpen our thinking and hypotheses.

What is the Role of Theory in the Life Sciences?

People say that to learn about the philosophy of science, one should not listen to what scientists say, but rather watch what they do. Most of the time, if cell biologists use theory at all, it appears at the end of their paper, a parting shot from figure 7. A model is proposed after the experiments are done, and victory is declared if the model 'fits' the data. But there is another way to go about using theory. This second approach not only provides a conceptual framework for experiments that have already been done but, more importantly, it also uses theory to produce interesting, testable predictions about experiments that have not yet been done. This type of theory often appears at the beginning of the paper, an opening volley from figure 1, to justify the experiments that follow. Here I describe the opportunity offered by practicing 'Figure 1 theory', where the theory comes first, and everything from the experimental design to the data analysis and interpretation flow from it.

It is an important time to reexamine the role of theory in biology. The explosion of data in the life sciences has created a deep tension between fact and concept. Indeed, the frenzy surrounding big data has led some to speculate 'the end of theory' [1]. The supposition is that if we can find the right correlations between different measurables, we need not bother with finding the underlying 'laws' that give rise to those correlations. The French mathematician Henri Poincaré famously noted 'A science is built up of facts as a house is built up of bricks. But a mere accumulation of facts is no more a science than a pile of bricks is a house'. Biology has many rooms and hallways of exquisite beauty, but there are still many bricks awaiting their place in the structure of biological science. Examples abound. Quantitative microscopy is now providing a picture of when and where the macromolecules of the cell are found. Mass spectrometry and fluorescence microscopy give an unprecedented look at the mean and variability in the number of mRNAs, lipids, proteins, and metabolites in cells of all kinds. DNA sequencing now routinely provides a base pair resolution view of genomes and their occupancy by proteins such as histones and transcription factors. Yet we are often lost amid the massive omic and imaging databases we have collected without a theoretical understanding to guide us. When successful,

[1]Department of Applied Physics and Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA

*Correspondence: phillips@pboc.caltech.edu (R. Phillips).

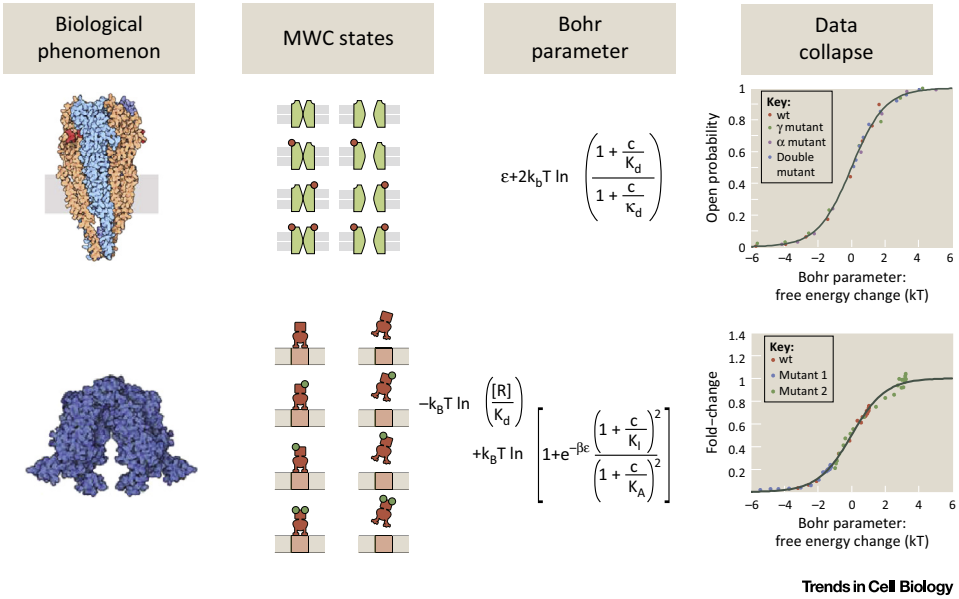

Figure 1. Broad Reach of Statistical Mechanical Models of Allostery. The top example shows an ion channel known as the nicotinic acetylcholine receptor and the bottom example shows the gene regulatory molecule known as Lac repressor. The Monod–Wyman–Changeux model (MWC) considers the inactive and active states in all of their different states of ligand occupancy [36]. The Bohr parameter provides the critical natural scaling variable that makes it possible for data from different mutants to all fall on one master curve as shown in the final column [27]. Different colored data points correspond to different mutants of the ion channel (top) or repressor molecule (bottom). Ion channel data from [37] and repressor data from [38].

Figure 1 theory tells us from the get-go exactly what data we need to collect to attempt to test our theoretical musings. As a result of the experimental advances driving cell biology, there is enormous pressure to turn facts into a corresponding conceptual picture of how cells work [2].

What exactly do we mean by theory? In many cases, our first understanding of some biological problem might be based on powerful, cartoon-level abstractions, already a useful first level of theory that can itself serve a Figure 1 role. These abstractions make qualitative predictions that we can then test. However, by mathematicizing these cartoon-level abstractions, we go farther, by formally committing to their underlying assumptions we can thus use the logical machinery of mathematics to sharpen our hypotheses and more deeply explore their consequences. Jeremy Gunawardena has amusingly but thoughtfully referred to this kind of theory as the exercise of converting our 'pathetic' thinking into mathematical form and then exploring the consequences of the assumptions behind that thinking [3].

How Can Theory Enlighten Us?

Where is the evidence that mathematical theory has the power to expand our understanding of the living world in the same way that microscopy, genetics, and biochemistry, for example, already have? In fact, as has been noted elsewhere, there is a long tradition of deep and fundamental biological insights that required quantitative analysis [3,4]. One of my personal favorites concerns the question of the physical limits on how cells can detect environmental stimuli. Quantitative reasoning has provided us with insights into processes as diverse as chemotaxis, in which cells can detect tiny chemical gradients, or vision, where networks of molecules make it possible for photoreceptors to detect small numbers of photons [5–7]. For example, in the context of chemotaxis, theoretical considerations shed deep light on the mechanisms of both gradient detection and how cells adapt to changes in the ambient chemoattractant concentration [5–8]. Another celebrated example is the way in which probability distributions serve as a window into biological mechanisms [9]. The famed Luria–Delbrück