Quantitative cell biology: the essential role of theory

Jonathon Howard
Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520

ABSTRACT
Quantitative biology is a hot area, as evidenced by the recent establishment of institutes, graduate programs, and conferences with that name. But what is quantitative biology? What should it be? And how can it contribute to solving the big questions in biology? The past decade has seen very rapid development of quantitative experimental techniques, especially at the single-molecule and single-cell levels. In this essay, I argue that quantitative biology is much more than just the quantitation of these experimental results. Instead, it should be the application of the scientific method by which measurement is directed toward testing theories. In this view, quantitative biology is the recognition that theory and models play critical roles in biology, as they do in physics and engineering. By tying together experiment and theory, quantitative biology promises a deeper understanding of underlying mechanisms, when the theory works, or to new discoveries, when it does not.

WHY QUANTITATE?
When I hear the word quantitation, what comes to mind is the contrast between two famous 19th-century German scientists, Alexander von Humboldt and Carl Gauss, as described in Measuring the World: A Novel (Kehlmann, 2009). von Humboldt was a great Prussian explorer who made many discoveries, including electric eels, in his expeditions to South and Central America, which included the Casiquiare Canal, which connects the Orinoco and Amazon Rivers. A man of extraordinary energy, he was the consummate quantitater—he measured the heights of mountains and the numbers of lice on the heads of natives. Gauss, as Surveyor and Astronomer for the Kingdom of Hannover, was also a quantifier, but with a different style. Charged with devising a practical method to survey on a spherical Earth, he laid the foundation for a school of geometry that culminated in the work of Riemann and the notion of curved space-time in Einstein’s general theory of relativity. What a contrast of styles!

The lesson here is that when you have a theory, measurement becomes very important and intellectually challenging. Have all the relevant parameters been measured? Have all the confounding complications been considered? Does the theory really fit to the data? How accurately have the underlying quantities (parameters) been measured? Measurement focuses the mind.

THEORIES, MODELS, AND SIMULATIONS
We need some definitions. What is a theory? What is a model? What is the difference? In biology, we usually distinguish three types of model: conceptual models, mathematical models, and simulations (Bowne-Anderson et al., 2013). They all serve the same general purpose: to explain more complex phenomena in terms of more elementary processes or molecular interactions. Models bridge scales.

A conceptual model is a proposed mechanism underlying experimental observations. This can be in the form of a cartoon or picture that conveys the author’s idea or explanation of his or her results. It often appears in “Figure 7”! The conceptual model is based on intuition and serves as a postulate. But can it really explain the data? Usually, more work is needed to establish the plausibility of a conceptual model. Perhaps the correct explanation is counterintuitive.

A mathematical model is the development of the conceptual model into a system of equations. The postulates are identified (often in the form of parameters), and the implications or predictions are derived by solving the equations. The equations usually describe the average behavior of a large number of interacting molecules or entities—a so-called mean-field theory (although variances and other statistical properties can often also be calculated). The solutions are then compared with the experimental measurements—the quantitation of the data. If the equations can be solved analytically, all is
well and good because this usually indicates that the solution is simple, or at least familiar. If not, numerical solutions are readily provided by such programs as MATLAB (www.mathworks.com). The predictions of the model (with appropriate parameters) can then be used to quantitatively account for the experimental data. This is the process of curve fitting, which can be used to ask whether the theory actually fits the data (i.e., establish plausibility) and to measure the underlying parameters and their uncertainties.

A simulation is a computer-based imitation of a system. Like a mathematical model, a simulation postulates interactions between underlying molecules or entities. Unlike a mathematical model, however, a simulation models individual behavior (rather than mean behavior) and often needs to be repeated a large number of times to obtain an accurate measure of the average behavior. Simulations are especially valuable when modeling small-number systems, which can display highly stochastic behavior. The simulation approach is facilitated by the ever-increasing speed of computers. Simulations can be used for establishing the plausibility of a conceptual model, and by changing the parameters, the simulator can experiment with the model to gain insight into how it works.

Simulations have two potential limitations. The first is that they often contain many details such as the precise initial conditions or the order of the interactions. In this sense they are often like experimental results, which also depend on many details (e.g., the composition of the buffers). In light of well-documented cases in which the results of simulations depended on these details, it has been argued that a simulation should be treated like an experimental result in the sense that it should be confirmed in different laboratories before being accepted (Mitchell, 2009).

A second potential limitation of simulation is that although it can establish plausibility of a mechanism, it might provide only limited insight. This is especially the case if the system behavior is unexpected and the elementary interactions give rise to so-called emergent or collective behavior. In this case, the behavior of the simulation may be as mysterious as the biological phenomenon being modeled, even when the simulator can probe the model by varying the model parameters (analogous to performing wet experiments). An alternative approach is to use so-called toy models (or idea models; Mitchell, 2009), which, while oversimplifying the system under study, are nevertheless simple enough to solve using a mathematical model, thereby providing insight into mechanism. An example of a toy model is Turing's reaction-diffusion mechanism of pattern formation (Turing, 1952; Gierer et al., 1972).

I do not want the reader to think that simulations are always inferior to mathematical models. Indeed, there are now very good algorithms that have solved many of the technical problems associated with simulations (e.g., the Gillespie algorithm; Doob, 1945; Gillespie, 2007). In addition, mathematical models have their own limitations. A good theorist can explain anything, given enough parameters! The physicist Freeman Dyson recounts attending a conference at which Enrico Fermi criticized the complexity of Dyson's model by quoting the mathematician John von Neumann (Dyson, 2004):

"With four parameters I can fit an elephant, and with five I can make him wiggle his trunk." Actually, it is not so easy to model an elephant with four parameters, although we recently achieved it, even making the trunk wiggle with the fifth parameter (Mayer et al., 2010a).

What is the difference between a model and a theory? If a model has a small number of premises or parameters that give rise to broad and/or unexpected implications, then we might call it a theory. Evolution—the modern evolutionary synthesis—is a theory because it makes a relatively small number of postulates yet predicts a very broad range of observations, including the laws of inheritance and models of population genetics. It could be argued that other models, especially those underlying spatial oscillations (e.g., Turing patterns) and temporal oscillations (e.g., the cell cycle; see later discussion), rise to the level of theories.

**EXAMPLES OF THE SUCCESSFUL USE OF MODELS IN CELL BIOLOGY**

Theories and models can be used to establish plausibility, sufficiency, or impossibility. They can also be used to measure.

The importance of establishing plausibility is illustrated by a recent example in which a textbook conceptual model was shown not to explain the observations. It was proposed that flow of the actin–myosin cortex was driven by a gradient of tension (Bray and White, 1988). However, using experiments (laser cutting) and theory, it was found that cortical flow in the one-cell *Caenorhabditis elegans* embryo is driven not by a gradient in tension but by a gradient of contractility (i.e., motor activity); in the direction of flow, there is no tension gradient, but, counterintuitively, in the transverse direction, there is a tension gradient but no flow (Mayer et al., 2010b)!

This example shows the importance of developing a mathematical model (and testing it) to establish the plausibility of a conceptual model.

Two examples of the use of theory to establish sufficiency are the action potential and the cell cycle. In a remarkable feat that serves as a foundation for neurobiology, Hodgkin and Huxley (1952) formulated and solved equations showing that the action potential in nerve cells can be accounted for by the voltage-dependent interactions between sodium and potassium conductances (channels in the modern parlance). The second example concerns the cell division cycle. Following seminal discoveries showing the necessity of phosphorylation for controlling the transitions between phases of the cell cycle, Tyson and Novak (2011) showed the sufficiency of phosphorylation networks to drive the transitions. In other words, they answered the question, "What do you need to build a molecular switch?" (answer: feedback and cooperativity), and, by doing so, they made the problem of the cell cycle "finite." In the process, they made key predictions that were confirmed experimentally and they explained many phenotypes.

An excellent example of the use of theory to show impossibility is the prescient work of Ken Machin (Machin, 1958). He proved that the periodic bending motion of cilia and flagella (such as sperm tails) cannot be driven by motor activity localized to the base of the cilium only: if the beating were driven by a whip-like process at the base, the amplitude would die out with a characteristic shape due to the damping from the fluid. This shape is not consistent with the observed beating patterns, whose amplitudes are maintained along the length of the flagellum. The conceptual model of a whip-like mechanism was therefore wrong. Instead, he proposed motors all along the length, with a traveling wave of activity propagating from base to tip. This was his discovery. Not bad, considering that the motor dynein had not been discovered (Gibbons and Rowe, 1965), nor was it known whether the filaments inside the flagellum (the microtubules) slid or contracted (they slide; Satir, 1965). Of interest, the traveling wave has still not been directly visualized. Showing impossibility is the most powerful use of mathematical models: it allows hypotheses to be falsified, which is difficult to achieve using simulations.

In my work, I have used theory as a tool to measure molecular properties and interactions. For example, we used precise mechanical measurements on hair cells—the sensory receptors underlying the sensation of sound and acceleration—to estimate the force needed to open a single channel, the displacement associated with the opening of the channel's gate, and the number of channels...
(Howard et al., 1988). In other work, we used theory to infer the bending stiffness of actin filaments and microtubules from an analysis of their shape fluctuations (Gittes et al., 1993). More recently, we showed how molecular interactions between kinesin motor proteins and microtubules can give rise to length-dependent microtubule depolymerization, which may play a key role in determining the size of organelles such as the mitotic spindle (Varga et al., 2006, 2009; Leduc et al., 2012).

WHERE TO, QUANTITATIVE BIOLOGY?
How are we to solve the important open questions in cell biology? These include the following:

1. Morphology: What determines the size, number, and shape of cells and organelles?
2. Signaling: How do molecules compute, and what is the role of spatial localization and compartmentalization?
3. Scaling: How do we bridge from one scale to another? Are there universal laws?

Especially important is to understand how variability is minimized within a cell to produce reproducible structures and signals, and amplified among cells to enlarge the number of stimuli to which a population of cells can respond.

Of course experimental tools are essential. There has been rapid development of single-molecule techniques, making it possible to manipulate individual molecules using optical and magnetic techniques and to visualize their activity inside cells using fluorescence microscopy. In this way molecular interactions can be studied in great detail. It is even possible to sequence single molecules of DNA (www.pacificbiosciences.com). At the whole-cell level the developments are as impressive: image processing techniques can be used to segment single cells in tissues to measure shape and activity, their total protein assayed by mass spectrometry, and their genome sequenced. These studies have revealed remarkable cell-to-cell variability of “identical” cells.

Hand in hand with these developments, we must have theory and models to make sense of the experimental results, as exemplified by Gauss, rather than the collection of data for their own sake. I predict that in the future, theory will drive biology as it does physics and engineering.

ACKNOWLEDGMENTS
I thank Hugo Bowne-Anderson and Marija Zanic for discussions on the different types of models, as well as for their comments on an earlier version of the article. I also thank Thierry Emonet, who has pioneered quantitative biology at Yale University and with whom I have benefited from many discussions.

REFERENCES
Bowne-Anderson H, Zanic M, Kauer M, Howard J (2013). Microtubule dynamic instability: a new model with coupled GTP hydrolysis and multistep catastrophe. Bioessays 35, 452–461.
Bray D, White JG (1988). Cortical flow in animal cells. Science 239, 883–888.
Dobel JL (1945). Markoff chains—denumerable case. Trans Am Math Soc 58, 455.
Dyson F (2004). A meeting with Enrico Fermi. Nature 427, 297–297.
Gibbons IR, Rowe AJ (1965). Dynein: a protein with adenosine triphosphatase activity from cilia. Science 149, 424–426.
Gierer A, Gierer A, Meinhardt H (1972). A theory of biological pattern formation. Kybernetik 12, 30–39.
Gillespie DT (2007). Stochastic simulation of chemical kinetics. Annu Rev Phys Chem 58, 35–55.
Gittes F, Mickey B, Nettleton J, Howard J (1993). The chemical basis of morphogenesis. Proc R Soc B Biol Sci 237, 37–72.
Hodgkin AL, Huxley AF (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117, 590–544.
Howard J, Hudspeth AJ, Hudspeth AJ (1988). Compliance of the hair bundle associated with gating of mechanoelectrical transduction channels in the bullfrog’s saccular hair cell. Neuron 1, 189–199.
Kehlmann D (2009). Measuring the World, New York: Vintage.
Leduc C, Padberg-Gehle K, Varga V, Helbing D, Diez S, Howard J (2012). Molecular crowding creates traffic jams of kinesin motors on microtubules. Proc Natl Acad Sci USA 109, 6100–6105.
Machin KE (1958). Wave propagation along flagella. J Exp Biol 35, 796–806.
Mayer M, Depken M, Bois JS, Jülicher F, Grill SW (2010b). Anisotropies in cortical tension reveal the physical basis of polarizing cortical flows. Nature 467, 617–621.
Mayer J, Khary K, Howard J (2010a). Drawing an elephant with four complex parameters. Am J Phys 78, 648–649.
Mitchell M (2009). Complexity Theory: A Guided Tour, Oxford, UK: Oxford University Press.
Sarit P (1965). Studies on cilia: II. Examination of the distal region of the ciliary shaft and the role of the filaments in motility. J Cell Biol 26, 805–834.
Turing A (1952). The chemical basis of morphogenesis. Proc R Soc B Biol Sci 237, 37–72.
Tyson CB, Novak B (2011). Irreversible transitions, bistability and checkpoint controls in the eukaryotic cell cycle: a systems-level understanding. In: Handbook of Systems Biology, ed. AJM Walhout, M Vidal, and J Dekker, Waltham, MA: Academic Press, 1–40.
Varga V, Helenius J, Tanaka K, Hyman AA, Tanaka TU, Howard J (2006). Yeast kinesin-8 depolymerizes microtubules in a length-dependent manner. Nat Cell Biol 8, 957–962.
Varga V, Leduc C, Bormuth V, Diez S, Howard J (2009). Kinesin-8 motors act cooperatively to mediate length-dependent microtubule depolymerization. Cell 138, 1174–1183.