REVIEW

Cell Polarity: Quantitative Modeling as a Tool in Cell Biology

Alex Mogilner,[1]* Jun Allard,[1] Roy Wollman[2]

Among a number of innovative approaches that have modernized cell biology, modeling has a prominent yet unusual place. One popular view is that we progress linearly, from conceptual to ever more detailed models. We review recent discoveries of cell polarity mechanisms, in which modeling played an important role, to demonstrate that the experiment-theory feedback loop requires diverse models characterized by varying levels of biological detail and mathematical complexity. We argue that a quantitative model is a tool that has to fit an experimental study, and the model's value should be judged not by how complex and detailed it is, but by what could be learned from it.

Four hundred years ago, Galileo observed that "Nature's great book is written in mathematical language." Since that time, physical phenomena have been described by mathematical equations, yet biology has remained qualitative. A possible explanation is that complex behavior in physics emerges from relatively simple interactions between many copies of few elements, whereas biological complexity results from nonlinear interactions of many heterogeneous species. In this sense, biological systems are similar to engineered machines (*1*): Inventories of both airplane parts and animal cell proteins consist of tens of thousands entries; cell interactomes look similar to machine blueprints; and performances of both engineering and biological structures are characterized by robustness and noise resistance (*2*). This analogy has limitations: Biological systems are built from stochastic and unreliable parts; are evolved rather than designed; and are subject to reverse, not direct, engineering. Nevertheless, in the last two decades, the mathematics usually applied to engineering and physics has been often used in cell biological studies where quantitative models serve as a guide for failing intuition (*3*).

The foundation for this surge was laid by two seminal papers that appeared 60 years ago. One was the biologically abstract and mathematically simple manuscript by Alan Turing (*4*) proposing that a pattern can emerge in an initially homogeneous mixture of two chemicals. Turing used two linear partial differential equations (PDEs) (Box 1) with few parameters to demonstrate that two chemicals, a slowly diffusing "activator" and a rapidly diffusing "inhibitor," could concentrate in different regions of space. Untested and unsubstantiated at the time, this conceptual model has served as a basis for many studies of polarity, chemotaxis, and development. Another work by Hodgkin and Huxley (*5*) was mathematically complex, grounded in experimental data and very detailed: Many ordinary differential equations (ODEs) (Box 1) with many parameters and nonlinearities were used to describe ion currents through voltage-gated channels in the axon membrane. The parameters and nonlinearities were measured, and the model reproduced the observed electric bursts in nerve cells, which revolutionized our understanding of excitable systems.

These two papers symbolize the opposite ends of "modeling space" (Fig. 1A). It is tempting to pronounce that we will be describing cells in ever more accurate terms and minute detail, moving from focused and conceptual (like ODEs describing three-node motifs in regulatory networks) to accurate and broad models (*6*), perhaps ending with a "whole-cell model" that completely recapitulates cell behavior on a computer, substitutes for wet laboratory experiments and makes personalized medicine possible. This is an appealing, if distant, goal. Meanwhile, this view subtly puts broad models above focused ones and suggests that there is a modeling "Road to Valhalla." Note, however, that our understanding of the nerve impulse progressed from the

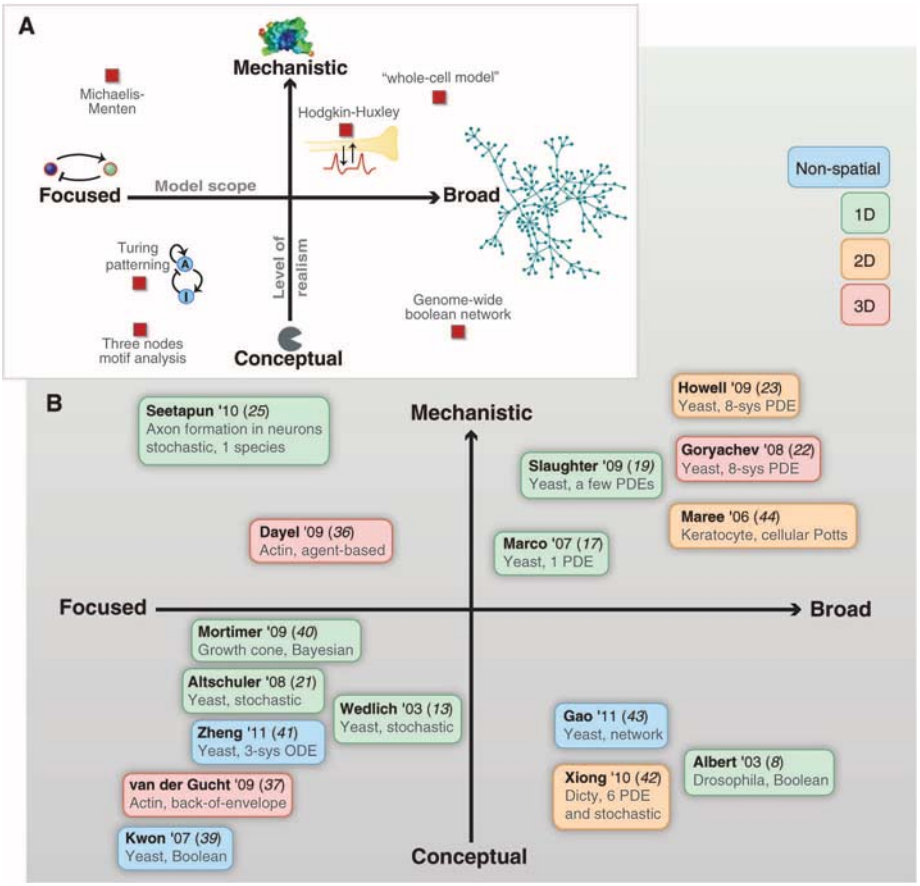

Fig. 1. Modeling space. (**A**) Computational models can be characterized by their scope and level of realism. Mathematically, focused models are simple, whereas broad ones are complex. Biologically, the models range from conceptual, offering mainly qualitative insight, to accurate and mechanistic, making many quantitative predictions. (**B**) Polarity models in modeling space marked by the first author, year in which the model was published, biological system the model applied to and mathematical method. Color corresponds to physical dimensionality of the model. Mathematical scope and level of realism do not correlate with dimensionality, type of math, or biological system.

[1]Department of Neurobiology, Physiology, and Behavior and Department of Mathematics, University of California, Davis, CA 95616, USA. [2]Department of Chemical and Systems Biology, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: mogilner@math.ucdavis.edu

Box 1. MODELING TOOLS

In mathematical biology, an **ODE** normally expresses the time derivative of a dynamic variable in terms of a function, usually nonlinear, of this variable. More frequently, a system of ODEs (dynamical system) appears. For example, if [*E*], [*S*], and [*C*] are concentrations of enzyme, substrate, and complex, respectively, then the dynamical system

$$\frac{d[E]}{dt} = -k_1[E][S] + k_2[C], \frac{d[S]}{dt} = -k_1[E][S], \frac{d[C]}{dt} = k_1[E][S] - k_2[C]$$

describes Michaelis-Menten kinetics. In this system, rates k_1 and k_2 are model parameters. ODEs are an extremely powerful tool as (i) vast mathematical apparatus (bifurcation diagrams and phase portraits, perturbation theory, and numerical analysis) has been developed to solve them; (ii) solving ODEs is almost trivial with much available user-friendly software; and (iii) their solutions provide very detailed mechanistic insight. The caveats of using ODEs for modeling are (i) chemicals have to be well mixed in the cell or their spatial gradients have to be neglected, (ii) detailed information about molecular interactions has to be available, and (iii) many data are necessary to validate the model.

A **PDE** is a generalization of an ODE: concentrations of chemicals (or distributions of molecular players, in general) can change not only in time, but in space as well, so not only time derivatives, but also (so-called "partial") derivatives with respect to spatial coordinates appear in the equations. For example, Michaelis-Menten kinetics in a 3D cell in the presence of diffusion will have the form:

$$\frac{\partial[E]}{\partial t} = -k_1[E][S] + k_2[C] + D_E\Delta[E], \frac{\partial[S]}{\partial t} = -k_1[E][S] + D_S\Delta[S],$$
$$\frac{\partial[C]}{\partial t} = k_1[E][S] - k_2[C] + D_C\Delta[C]$$

where D's are diffusion coefficients, and $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. These PDEs constitute a so-called reaction-diffusion system. Instead of diffusion, other transport processes (e.g., directed motor-driven transport) can be represented mathematically as well. PDEs have to be accompanied by boundary conditions that have to be chosen carefully. Similar to ODEs, PDEs are great for mechanistic insight. Solving PDEs is far from trivial. ODEs' caveats (ii) and (iii) apply to PDEs as well.

Unlike deterministic models, **stochastic simulations** usually describe molecules not as concentrations, but as random numbers. These numbers change in time with certain probabilities that are functions of random variables (numbers of these and other molecules) in the system. Many different types of stochastic models exist, for example: (i) direct Monte Carlo simulations, in which the computer generates random changes in the random variables at equal time increments; (ii) Gillespie simulations, in which computer calculates random time intervals at which the state of the system changes; (iii) Langevin equations that allow the addition of random steps to deterministic changes; (iv) Smoluchowski equations that introduce a probability density for system states and transform random equations into deterministic PDEs, and so on. The great advantage of stochastic simulations is that they expose the effects of fluctuations and noise that are often enormous in cells because of the relatively small number of molecules involved. The caveat is that usually stochastic simulations produce but a single random trajectory of the system (Smoluchowski equations are a notable exception, but these are often very hard to solve), so multiple, computationally expensive simulation runs and nontrivial statistical tools are needed for thorough investigation.

Agent-based simulations that are sometime deterministic, but more often stochastic, rely on explicitly tracking all essential molecules so that, on the basis of the current positions and states of these molecules, all interactions (chemical and mechanical) are computed and movements and states of every molecule are calculated based on the rules of physical chemistry and classical mechanics. Usually, thousands of such molecule-agents are simulated, the numerical codes are very involved, and computational expense is enormous. Often, these simulations suffer from limited qualitative insight—in a way, they substitute a computational "black box" for the live cell—but these problems are usually more than offset by two benefits: life-like resulting movies that can be compared directly with time-lapse microscopy and the ability to perform lifelike perturbation experiments on the computer. Other useful modeling techniques, such as cellular automata and Potts models, are midway between PDEs and agent-based simulations.

Boolean networks are well suited to reproduce the qualitative behavior of extensive networks when the amount of quantitative experimental data is limited. In a Boolean representation, the biological active or inactive state of a species are represented by the on-off states of nodes in an "interactome" network. Logical rules based on qualitative insight prescribe switches between states of each node, depending on the state of the nodes to which the node is connected in the network. Boolean models are useful for a fast quantitative look at the dynamics of large biological networks, but because of the difficulties of treating physical time and some mathematical artifacts, such models are gradually falling out of favor.

Bayesian models use the rules of probability theory (Bayes' formula for conditional probabilities) and experimental multivariate data that depict causal relationships between biological variables to uncover statistical (and/or causal) relations among these variables. These models lack a time dimension and therefore cannot include feedback loops that are prevalent in biology.

Network analysis uses the sophisticated methods of graph theory, topology, statistics, and combinatorics to find modules, motifs, and other building blocks representing small standard dynamic systems in large biological networks. The methods of network analysis are extremely useful for large networks, and relevant software is becoming better and more widely available.

complex Hodgkin-Huxley model through the FitzHugh-Nagumo model (*7*) that reduced the system radically to two PDEs and few parameters. Although simplistic, the FitzHugh-Nagumo model was amenable to deeper mathematical analysis, by building intuition that could be applied to full biophysical reality. Besides, increasing a model's realism does not necessarily mean more difficult math: There are simple and accurate models, like Michaelis-Menten kinetics equations (Fig. 1A) (Box 1). There are also broad and detailed, yet conceptual models, such as the Boolean description (Fig. 1A) (Box 1) of segment polarity gene expression (*8*). Here, we review recent studies of budding yeast to illustrate how a variety of modeling approaches advanced our understanding of cell polarity (Fig. 2). We argue that quantitative modeling is a versatile tool that has to fit the biological problem and can be judged by its usefulness rather than its comprehensiveness or sophistication.

Modeling Yeast Polarity

Many cells polarize in response to external cues such as preexisting landmarks, chemoattractants, or contact with other cells (*9*). In budding yeast cells, cortical landmark proteins inherited from previous division cause localized activation of the

guanosine triphosphatase (GTPase) Rsr1, leading to recruitment of the polarization machinery components, including the key GTPase Cdc42 (Fig. 2A) and, ultimately, to symmetry breaking and switching from isotropic growth to growth along a polarized axis (*10*). However, landmark proteins are not essential for the emergence of the Cdc42 cluster that marks future growth site: When the landmark is bypassed by Rsr1 removal or constitutive activation of $Cdc42^{Q61L}$ (in which glutamine at position 61 is replaced by leucine and which cannot hydrolyze guanosine triphosphate and respond to upstream signals), clusters of concentrated Cdc42 still form, albeit at random locations, so the cell is able to self-polarize (Fig. 2A). This phenomenon is arguably the best understood paradigm of cell polarity, in no small measure owing to the power of yeast genetics and many studies that judiciously combined microscopy and modeling to allow detailed mechanistic insights.

Turing's pioneering paper and several models it inspired (*11*, *12*) predict that locally acting positive-feedback loops and globally acting negative regulators can lead to self-polarization. [The models we review below share the general philosophy of Turing's mechanism, but in a narrower, mathematical sense they differ from Turing system; see (*12*) for details.] Immediate questions arise from this prediction: Is it indeed the Turing mechanism that is responsible for self-polarization? What are the molecular identities of the activators and inhibitors? What are the transport mechanisms underlying the process? How are complex, specific combinations of regulatory pathways wired to achieve polarity?

Less than a decade ago, Wedlich-Soldner *et al.* (*13*) started to answer these questions quantitatively by using mutants to establish that actin-myosin–directed vesicle transport and fusion were essential for $Cdc42^{Q61L}$ polarization. They proposed that Cdc42 molecules would be recruited to a localized cap on the membrane through transport along actin cables, whereas deposition of Cdc42 to the cap would stimulate further actin accumulation at this site (Fig. 2B). As it was hard to prove experimentally that such a positive-feedback circuit is sufficient to induce polarization, Wedlich-Soldner *et al.* performed simple stochastic simulations (Box 1) in which actin cables nucleated at a rate proportional to the local membrane Cdc42 concentration, Cdc42 was delivered along each cable at a rate proportional to Cdc42 cytoplasmic concentration, and Cdc42 settled on the plasma membrane in a bell-shaped distribution around the cable. Multiple caps were often observed in a single cell, and the model predicted, correctly, that the cap number increases with initial membrane Cdc42 concentration.

Two years later, Ozbudak *et al.* (*14*) studied the dynamics of this symmetry breaking with Rsr1 deleted, rather than by expressing higher levels of activated Cdc42. To their surprise, they found that the single Cdc42 peak moved around the cell. When actin was inhibited, the Cdc42 peak remained, but its movement ceased. To explain this phenomenon, Ozbudak *et al.* turned to the experimental finding that the scaffold protein Bem1 is an essential component in an actin-independent self-polarization pathway (*15*) and to the theoretical finding that certain negative-feedback loops can result in traveling waves (*16*). They assumed that actin cables deliver molecules initiating an actin patch on the membrane causing a dispersal, rather than concentration of Cdc42, which ultimately means that Cdc42 molecules on the membrane inhibit further local accumulation of Cdc42 with a time delay dependent on the actin dynamics. A PDE with respective negative feedback and a time-delayed term, in addition to a hypothetical autocatalytic reaction term describing Bem1-dependent accumulation of Cdc42 on the membrane and Cdc42 diffusion term, mathematically explained the traveling wave. This conceptual model, besides making sense of the data, proposed that actin dynamics are part of a negative-, not positive-, feedback loop.

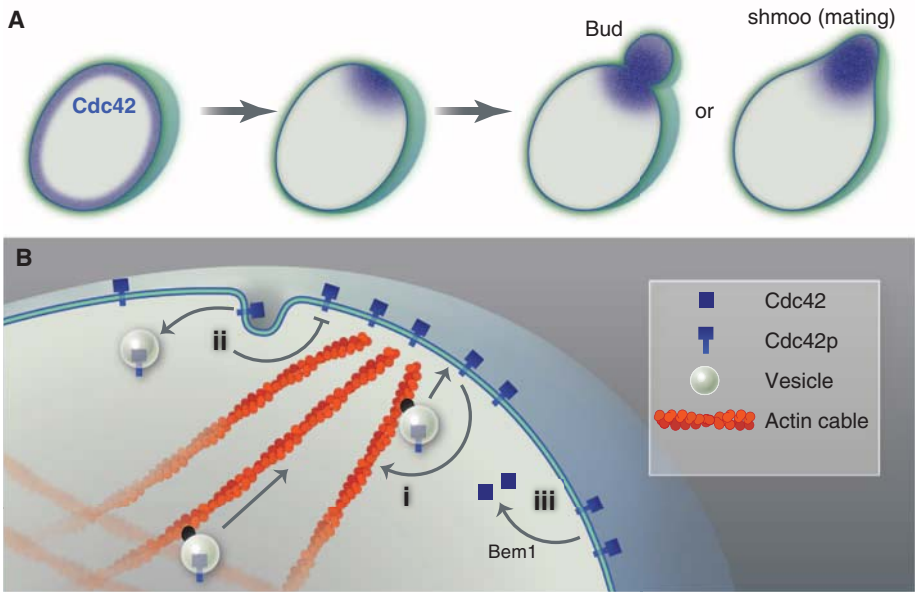

Fig. 2. Budding yeast polarization. (**A**) Initially symmetric cell with uniform distribution of Cdc42 becomes polarized, either in response to an extracellular cue or spontaneously. Polarization of Cdc42 distribution leads to polarized growth, either for budding or mating. (**B**) Feedback loops leading to Cdc42 polarization: (i) positive, actin-dependent feedback, (ii) endocytosis, and (iii) positive feedback via autocatalytic Cdc42 membrane recruitment mediated by Bem1.

Marco *et al.* proposed an alternative solution (*17*) based on data showing that endocytosis and lateral diffusion in the membrane are essential for yeast polarization (*18*). They suggested that a combination of three processes—actin-Cdc42–positive accumulation, Cdc42 lateral diffusion, and removal by endocytosis, together, could maintain the polarized state (Fig. 2B). By combining three respective mathematical terms in a simple PDE and solving this equation, Marco *et al.* found that both emergence and maintenance of the polarization cap could be explained. The model made a nontrivial prediction: There is an optimal rate of endocytosis in terms of how "sharp" the Cdc42 polarization is, because slower endocytosis causes spreading of the Cdc42 cap over the surface, whereas faster endocytosis depletes the Cdc42 cluster. Measurements of the rates of transport, endocytosis, and geometric parameters indeed resulted in the predicted optimal rate. Compared with the previous model, this one was more accurate and detailed, with parameters fitted from the data.

The model of Marco *et al.* was based on the measurements of a mutant, $Cdc42^{Q61L}$, which is stably associated with the membrane. A couple of years later, Slaughter *et al.* (*19*) took their model to the next level of accuracy by trying to explain how cells maintain the dynamic distribution of the wild-type Cdc42, which transitions between the membrane and cytoplasm at higher rates than $Cdc42^{Q61L}$. Slaughter *et al.* based their model on data indicating that actin-dependent and independent pathways play redundant but essential roles in maintaining Cdc42 polarization. By assuming that these two pathways work in parallel to control Cdc42 recycling at the polar cap and by adding respective mathematical terms [making the model in (*17*) more detailed and precise], Slaughter *et al.* found that, depending on the data-supported model parameters, the shape of the Cdc42 peak resembles either a bud that the cell grows to enter the mitotic cycle or a shmoo that grows as a mating projection (Fig. 2A). Thus, the model made a provocative prediction that parameters of Cdc42 recycling in yeast are adapted not to achieve maximum polarity but to fulfill specific morphogenetic outcomes.

A recent study (*20*) carefully examined the assumptions used in (*17*) and noticed that the

previous model treated Cdc42 traffic to and from the polarization cap as a direct protein flux, without taking into account the membranes that actually transport the Cdc42. However, if the membrane flux is taken into account, then, in the steady state, the amount of membrane in vesicles undergoing endo- and exocytosis with the cap have to be the same, so Cdc42 concentrations in the endo- and exocytic vesicles have to be the same. The problem is that, if the Cdc42 concentrations in the endo- and exocytic vesicles are equal to local concentrations in the cap and cytoplasm, respectively, then higher Cdc42 concentration in the cap would cause faster Cdc42 endocytosis and, ultimately, depolarization. In fact, new data and mathematical analysis in (*20*) suggested that Cdc42 would have to be significantly and unevenly concentrated into the endo- and exocytic vesicles for the model of (*17*) to work. Results of (*20*) will undoubtedly stimulate future studies to determine whether such traffic mechanisms exist.

Altschuler *et al.* (*21*) reversed the apparent trend toward more accurate and detailed models. They investigated the actin-independent polarization pathway using a simple, stochastic—rather than deterministic—model in which Cdc42 molecules on the membrane catalyzed recruitment to the membrane of cytoplasmic Cdc42. They found that, if the total number of Cdc42 molecules is small, then the stochastic effect of one emerging Cdc42 cluster "grabbing" a majority of signaling molecules leads to self-polarization, whereas greater Cdc42 numbers predicted global and homogeneous Cdc42 membrane recruitment. The data indeed showed that the frequency of polarization decreases as the number of molecules becomes large. This conceptual model illustrated the role of stochastic effects, fundamental for cell biological processes in which the number of molecules involved is often small.

In the same year, another mathematical model investigated the actin-independent pathway (*22*) on the basis of a very different philosophy. All previous models were top-down, based on coupling "modules," with molecular details left to be clarified later, by assumed nonlinear interactions. In contrast, Goryachev and Pokhilko (*22*) built a bottom-up model by describing simple mass-action reactions and diffusion for all known components of the actin-independent pathway in mathematical terms, solving respective equations and confirming that this fine-grained model predicts self-polarization without any additional assumptions. Then, using network analysis (Box 1), Goryachev and Pokhilko found a motif in the large signaling network responsible for the polarization instability. They predicted that competition of the Cdc42 clusters in the membrane for the limited cytoplasmic pool of rapidly diffusing Bem1-containing complexes is at the core of this motif. Hence, two separate modeling approaches, detailed and schematic, used in the same paper, both demonstrated the model's feasibility and built intuition.

All these models inspired a recent powerful study (*23*) in which Howell *et al.* tested the redundant polarization mechanisms by creating a fusion protein that effectively tethered Bem1 to the membrane. This effectively weakened the diffusion-mediated mechanism and validated actin-mediated positive feedback. They noted that the synthetically rewired cells often polarized to two sites simultaneously. Combined experimental-theoretical analysis of both wild-type and rewired cells led to the understanding that yeast cells polarize to a single "front" because of competition of membrane Cdc42p clusters for a limiting pool of Bem1-Cdc42 complexes. If such competition is slow, as in rewired cells, two buds could form.

A power of conceptual modeling is that ideas that arise in one model are often applicable to other phenomena. Polarization plays an important role in a wide variety of systems from neurons (Fig. 3A) (*24–26*) to *Caenorhabditis elegans* development (Fig. 3D) (*27–29*) to cell migration (Fig. 3B) (*30–34*) to mechanical symmetry breaking in actin gels (Fig. 3C) (*35–38*). In all these cases, a critical role in the establishment of polarity is played by the intricate interplay of positive- and negative-feedback loops, understanding of which is impossible without modeling.

Summary

Much progress has been made in understanding cell polarity by using models that not only summarized experimental findings but inspired further experiments by forcing researchers to think rigorously about what can be assumed and motivating more accurate observations. As the history of the yeast polarization modeling illustrates, modelers did not simply increase the models'

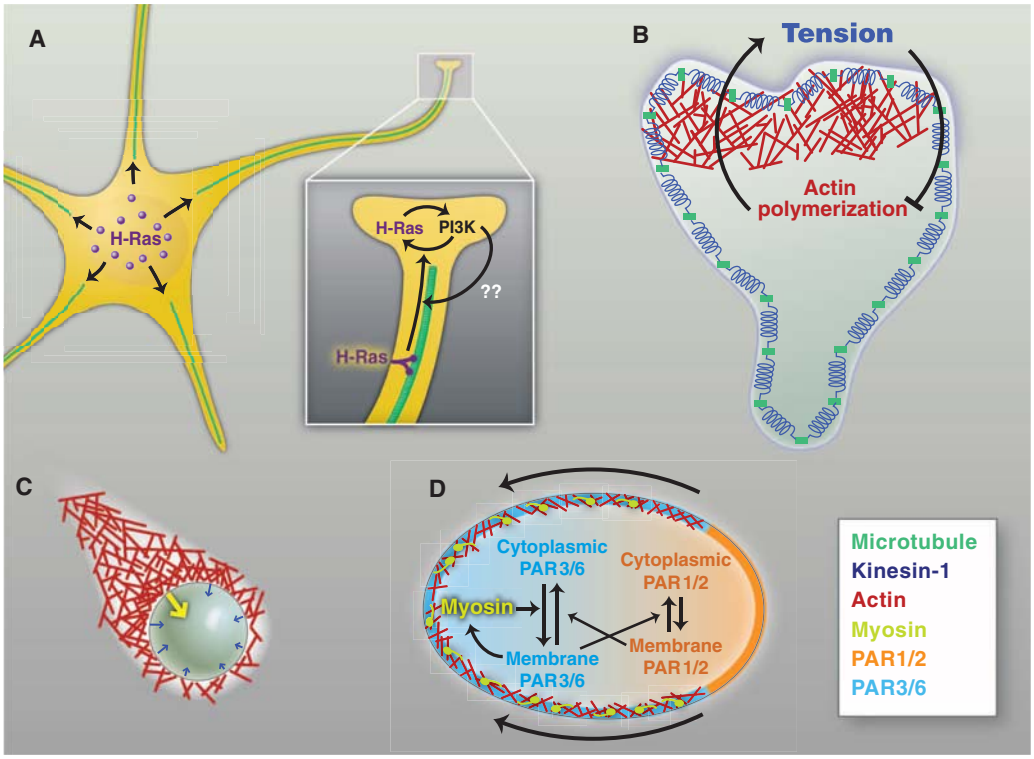

Fig. 3. Models of polarization in nonyeast systems. (**A**) Neuronal protrusions compete for limited pool of H-Ras. This competition acquires "winner-takes-all" features because of a local positive feedback between H-Ras and phosphatidylinositol 3-kinase (PI3K) at the protrusion tips and increased microtubule-based transport in the nascent axon. (**B**) Polarization of cell motility initiation is based on local actin polymerization that generates membrane tension, which propagates and inhibits actin polymerization globally. Myosin is an additional, local, inhibitor of the actin polymerization. (**C**) Symmetry breaking of actin gel that grows around a bead is driven by positive mechanical feedback (yellow arrow) between pushing actin filaments at denser actin arrays. Uniform pushing and tension around the bead (blue arrows) play the role of global inhibition. Autocatalytic breaking of the actin gel can also be a local activator. (**D**) Mechanochemical polarization of PAR proteins in *C. elegans*. Polarization is based on mutual inhibition of PAR proteins via their transitions between the cortex and the cytoplasm and myosin-dependent transient cortical flow (outside arrows).

complexity but rather moved nimbly within quadrants of the two-dimensional (2D) modeling space charted by two orthogonal axes characterizing model scope and level of realism (Fig. 1B).

A look at this space illustrates that to achieve qualitative insight, simple Boolean (*39*), Bayesian (*40*) (Box 1), ODE (*41*), or stochastic (*13*, *21*) models or physical estimates (*37*) that are focused and conceptual may include few details and make few predictions, but these predictions can be important. A model can be "bigger" and its scope more broad, but the level of realism can stay similar to that of the focused conceptual models. For example, broad and conceptual PDE (*42*), Boolean (*8*), or network (*43*) models can describe mathematically very large interacting systems but use only causal links between genes and/or proteins and so predict just qualitative features of emergent spatial-temporal patterns. However, focused models with relatively few mathematical details (*25*, *36*) can be accurate and mechanistic when precise numbers matter as well as qualitative insight. The broad and mechanistic models (*17*, *19*, *22*, *23*, *44*) are useful when there is a need to mathematically integrate detailed quantitative data and to test precisely formulated hypotheses.

Outlook

Cell biology is transitioning into a quantitative science characterized by increasing integration of modeling into experiment. In this transition, we have to proceed with numerous, often arbitrary, assumptions about the nature of processes and parameter values governing cell systems. One great future challenge is to improve quantitative experimental methods with an eye toward synchronizing modeling and experiments. Then, frequent back-and-forth between theory and experiment using models of varying scope and level of realism will allow us to overcome the arbitrariness and uncertainty. Another significant challenge is to make switching from one type of model to another a more standard, less ad hoc procedure, to ease modeling use and integration between theory and experiment. Models along this course should be considered impermanent and should be judged by how useful they are and what we can learn from them, not by how close we are to the elusive whole-cell model.

References and Notes

1. G. T. Reeves, S. E. Fraser, *PLoS Biol.* **7**, e21 (2009).
2. A. D. Lander, *Cell* **144**, 955 (2011).
3. J. M. G. Vilar, C. C. Guet, S. Leibler, *J. Cell Biol.* **161**, 471 (2003).
4. A. M. Turing, *Philos. Trans. R. Soc. Lond. B* **237**, 37 (1952).
5. A. L. Hodgkin, A. F. Huxley, *J. Physiol.* **117**, 500 (1952).
6. S. J. Vayttaden, S. M. Ajay, U. S. Bhalla, *ChemBioChem* **5**, 1365 (2004).
7. R. FitzHugh, *Biophys. J.* **1**, 445 (1961).
8. R. Albert, H. G. Othmer, *J. Theor. Biol.* **223**, 1 (2003).
9. W. J. Nelson, *Nature* **422**, 766 (2003).
10. J. M. Johnson, M. Jin, D. J. Lew, *Curr. Opin. Genet. Dev.* **21**, 740 (2011).
11. M. D. Onsum, C. V. Rao, *Curr. Opin. Cell Biol.* **21**, 74 (2009).
12. A. Jilkine, L. Edelstein-Keshet, *PLOS Comput. Biol.* **7**, e1001121 (2011).
13. R. Wedlich-Soldner, S. Altschuler, L. Wu, R. Li, *Science* **299**, 1231 (2003).
14. E. M. Ozbudak, A. Becskei, A. van Oudenaarden, *Dev. Cell* **9**, 565 (2005).
15. J. E. Irazoqui, A. S. Gladfelter, D. J. Lew, *Nat. Cell Biol.* **5**, 1062 (2003).
16. H. Meinhardt, *J. Cell Sci.* **112**, 2867 (1999).
17. E. Marco, R. Wedlich-Soldner, R. Li, S. J. Altschuler, L. F. Wu, *Cell* **129**, 411 (2007).
18. J. Valdez-Taubas, H. R. B. Pelham, *Curr. Biol.* **13**, 1636 (2003).
19. B. D. Slaughter, A. Das, J. W. Schwartz, B. Rubinstein, R. Li, *Dev. Cell* **17**, 823 (2009).
20. A. T. Layton *et al.*, *Curr. Biol.* **21**, 184 (2011).
21. S. J. Altschuler, S. B. Angenent, Y. Wang, L. F. Wu, *Nature* **454**, 886 (2008).
22. A. B. Goryachev, A. V. Pokhilko, *FEBS Lett.* **582**, 1437 (2008).
23. A. S. Howell *et al.*, *Cell* **139**, 731 (2009).
24. M. Fivaz, S. Bandara, T. Inoue, T. Meyer, *Curr. Biol.* **18**, 44 (2008).
25. D. Seetapun, D. J. Odde, *Curr. Biol.* **20**, 979 (2010).
26. N. Inagaki, M. Toriyama, Y. Sakumura, *Dev. Neurobiol.* **71**, 584 (2011).
27. N. W. Goehring *et al.*, *Science* **334**, 1137 (2011).
28. F. Tostevin, M. Howard, *Biophys. J.* **95**, 4512 (2008).
29. A. T. Dawes, E. M. Munro, *Biophys. J.* **101**, 1412 (2011).
30. M. M. Kozlov, A. Mogilner, *Biophys. J.* **93**, 3811 (2007).
31. F. Ziebert, S. Swaminathan, I. S. Aranson, *J. R. Soc. Interface* (2011).
32. D. Shao, W.-J. Rappel, H. Levine, *Phys. Rev. Lett.* **105**, 108104 (2010).
33. D. Kabaso, R. Shlomovitz, K. Schloen, T. Stradal, N. S. Gov, *PLOS Comput. Biol.* **7**, e1001127 (2011).
34. A. R. Houk *et al.*, *Cell* **148**, 175 (2012).
35. K. Sekimoto, J. Prost, F. Jülicher, H. Boukellal, A. Bernheim-Grosswasser, *Eur. Phys. J. E* **13**, 247 (2004).
36. M. J. Dayel *et al.*, *PLoS Biol.* **7**, e1000201 (2009).
37. J. van der Gucht, C. Sykes, *Cold Spring Harb. Perspect. Biol.* **1**, a001909 (2009).
38. J. S. Bois, F. Jülicher, S. W. Grill, *Phys. Rev. Lett.* **106**, 028103 (2011).
39. Y.-K. Kwon, K.-H. Cho, *Biophys. J.* **92**, 2975 (2007).
40. D. Mortimer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10296 (2009).
41. Z. Zheng, C.-S. Chou, T.-M. Yi, Q. Nie, *Math. Biosci. Eng.* **8**, 1135 (2011).
42. Y. Xiong, C.-H. Huang, P. A. Iglesias, P. N. Devreotes, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17079 (2010).
43. J. T. Gao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7647 (2011).
44. A. F. M. Marée, A. Jilkine, A. Dawes, V. A. Grieneisen, L. Edelstein-Keshet, *Bull. Math. Biol.* **68**, 1169 (2006).

Acknowledgments: We thank R. Li and D. Lew for helpful discussions. This work was supported by NIH grant 2R01GM068952 and NSF grant DMS-1118206 to A.M.

10.1126/science.1216380

REVIEW

Integrating Genomes

D. R. Zerbino,[1] B. Paten,[1] D. Haussler[1,2]*

As genomic sequencing projects attempt ever more ambitious integration of genetic, molecular, and phenotypic information, a specialization of genomics has emerged, embodied in the subdiscipline of computational genomics. Models inherited from population genetics, phylogenetics, and human disease genetics merge with those from graph theory, statistics, signal processing, and computer science to provide a rich quantitative foundation for genomics that can only be realized with the aid of a computer. Unleashed on a rapidly increasing sample of the planet's 10^{30} organisms, these analyses will have an impact on diverse fields of science while providing an extraordinary new window into the story of life.

Since the first genome sequences were obtained in the mid-1970s (*1*, *2*), computers have been necessary for processing (*3*) and archiving (*2*, *4*) sequence data. However, the discipline of computational genomics traces its roots to 1980, when Smith and Waterman developed an algorithm to rapidly find the optimal comparison (alignment) of two sequences of length n among the more than 3^n possibilities (*2*, *5*), and Stormo *et al.* built a linear threshold function to search a library of 78,000 nucleotides of *Escherichia coli* messenger RNA sequence for ribosome binding sites (*6*). What seemed large data sets for biology then don't seem so today, as high-throughput, short-read sequencing machines churn out terabytes of data (*2*, *7*). We have seen a 10,000-fold sequencing performance improvement in the past 8 years, far outpacing the estimated 16-fold improvement in computational power under Moore's law (*8*). Using genomics data to model genome evolution, mechanism, and function is now the heart of a lively field.

Every genome is the result of a mostly shared, but partly unique, 3.8-billion-year evolutionary journey from the origin of life. Diversity is created mostly by copy errors during replication. These create single-base changes, which are known as substitutions if spread to the whole population (fixed) or single-nucleotide polymorphisms (SNPs) if not uniformly present in the population (segregating). Replication errors also create insertions and deletions (collectively, indels), as well as tandem duplications where a short sequence is repeated sequentially. Chromosomes often exchange long similar segments through the process of homologous recombination. Specific sequences of DNA, known as transposable elements, have the

[1]Center for Biomolecular Sciences and Engineering, University of California, Santa Cruz, CA 95064, USA. [2]Howard Hughes Medical Institute, University of California, Santa Cruz, CA 95064, USA.

*To whom correspondence should be addressed. E-mail: haussler@soe.ucsc.edu