Developmental Cell *11*, 279–287, September, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.devcel.2006.08.004

Quantitative Modeling in Cell Biology: What Is It Good for?

Review

Alex Mogilner,[1,*] Roy Wollman,[2] and Wallace F. Marshall[3]
[1] Department of Neurobiology, Physiology and Behavior, and Department of Mathematics
[2] Department of Molecular and Cell Biology and Center for Genetics and Development
University of California, Davis
Davis, California 95616
[3] Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143

Recently, there has been a surge in the number of pioneering studies combining experiments with quantitative modeling to explain both relatively simple modules of molecular machinery of the cell and to achieve system-level understanding of cellular networks. Here we discuss the utility and methods of modeling and review several current models of cell signaling, cytoskeletal self-organization, nuclear transport, and the cell cycle. We discuss successes of and barriers to modeling in cell biology and its future directions, and we argue, using the field of bacterial chemotaxis as an example, that the closer the complete systematic understanding of cell behavior is, the more important modeling becomes and the more experiment and theory merge.

Why Does Cell Biology Need Quantitative Models?

Many individual chemical reactions in the cell involve a single enzyme catalyzing a single, well-defined transition in a substrate. Such reactions produce simple behaviors that can be plotted with a single exponential curve and understood with just a couple of Michaelis-Menten constants. In such cases, it is simple to go from experimental measurements to theoretical understanding, because there are few underlying players, and the overall behavior occurs on few time scales and can be represented by so few parameters that you can count them on one hand (Segel, 1988).

Physics and chemistry has shown us, however, that even when but a few players interact in nonlinear ways, complex and startling large-scale phenomena can result. For example, in the Belousov-Zhabotinskii reaction, a simple mixture of just a few chemicals spontaneously forms pulsating spiral waves in the beaker (Tyson, 1976). Physics, chemistry, and mathematics have also taught us that while large-scale behavior is often counterintuitive and surprising, it is not necessarily unpredictable, provided one employs an appropriate quantitative model in the form of a set of mathematical equations or a computer code. The model is successful if observed behaviors emerge from qualitative or quantitative analysis of these equations/codes. For instance, by using dynamical systems equations, it is possible to explain oscillations in Belousov-Zhabotinskii reaction (Field and Noyes, 1974). In a way, mathematics can be thought of as a crutch for our intuition, to help us bridge the gap between what we can see and what we can think about.

The exact tools that have been used previously in physical and chemical cases may simply not be directly applicable to cell biology. This has nothing to do with vitalism, but simply with the fact that cell biology occurs on multiple and radically different scales, both in terms of time, space, and complexity. Understanding how cellular-scale behaviors arise from molecular actions is fundamentally difficult due to the large number of many different kinds of molecules all interacting in complex networks. Another difficulty is that cell biological systems consist of thousands of molecules and so are not microscopic, but they are not macroscopic either; often, fluctuations of chemical or physical quantities in the cell are comparable in magnitude to the average values of these quantities.

The main difference between modeling in biology and physics stems from the inherent redundancy and heterogeneity of evolved molecular machines that have to be elucidated by "reverse engineering." This makes cell modeling very difficult but also unavoidable as new technologies produce staggering amounts of data about the spatiotemporal behavior of molecular assemblies. With increasing frequency, these data are quantitative—correlation functions, statistical regressions, and other similarly sophisticated forms—that cannot be reduced to simple qualitative statements, so mere qualitative "cartoon drawing" in the discussion section of a paper is not sufficient. Rather, to integrate and make sense of these data, quantitative modeling is needed as a "hypotheses generating machine" and a natural "endpoint" for the experimental efforts.

The most critical issue in developing a model is deciding its scale. For instance, to capture the biochemistry underlying the dynamic instability of microtubules, one could model the movements and interactions of each individual tubulin dimer as they exchange on and off the ends and undergo GTPase reactions (VanBuren et al., 2005). Alternatively, one could simply propose stochastic rules describing how the length of the microtubule is expected to change over time (Gliksman et al., 1993). In between these two extremes, there is still tremendous latitude in deciding how fine-grained a model to employ. Which approach is right?

Choosing between a fine-grained and a coarse-grained model involves several tradeoffs. On the one hand, constructing a coarser-grained model will often rely on extensive prior intuition about the cellular phenomenon. On the other hand, a finer-grained model could require more detailed prior information about the properties of the individual components. Beyond the simple pragmatic issues of which type of model is easier to construct, there is a much more important issue concerning how the model will be used.

A quantitative model can be exploited in many ways. First, the mere fact that such a model can recapitulate

*Correspondence: mogilner@math.ucdavis.edu

the phenomenon of interest is exceedingly important, because it indicates that we have identified all of the necessary players and interactions. On the other hand, a negative result is also important, indicating that we probably do not qualitatively understand a molecular mechanism. Second, models can show us which aspects of the system are critically important and which ones matter less. One can run the simulation with a range of values for some parameter of interest and thereby learn whether that parameter plays a key role or not.

Finally, the choice of modeling methods depends on the scale and purpose of the model. For example, statistical methods of bioinformatics are useful for identifying promoter binding sites; protein interaction network models are good for identifying all the genes involved in a certain function. However, these approaches would have a hard time answering the question: how does the molecular machine really work; e.g., what molecular processes underlie dynamic instability of microtubules? Listing the genes involved in this process or drawing a network of their interactions will not tell why microtubules undergo alternate catastrophes and rescues. Low dimensional differential equation models are appropriate to answer one part of this question and to understand how simple individual molecular machines that constitute cellular networks work. To answer another part of the question and to unravel organization of complex cellular networks, one has to use large-scale computational models.

Here we review some recent cell biological models. We narrowed our choices to explicit, dynamic, mechanistic, bottom-up models that make specific assumptions about pathways and the physics of biological phenomena. Thus, we left out wonderful top-down models that examine global properties of biological networks (Joyce and Palsson, 2006), identify network motifs and explore network stability and robustness (Prill et al., 2005; Yeger-Lotem et al., 2004), as well as integrative efforts that strive to model the entire cell (Ma'ayan et al., 2005). We also did not mention powerful causal and informatics models, both small- (Weinreb et al., 2006) and large-scale (Janes et al., 2006). What is exciting about the growing popularity of modeling in cell biology is the use of models as sophisticated working hypotheses helping to understand and design experiments, so the main requirement in our selection was a clear connection between theory and experiment. Even with these constraints, considering the recent and ongoing surge in the number of prominent joint experimental/modeling studies, we had to ignore important examples, for instance, the recent elegant unraveling of mechanisms of eukaryotic cell chemotaxis (Janetopoulos et al., 2004; Schneider and Haugh, 2005) and many others. We use the examples to address the fundamental questions that are normally not elaborated on in research papers:

What exactly does modeling add to experimental studies?
What is the appropriate model scale?
What are the appropriate modeling methods?

Further discussion of quantitative modeling in cell biology can be found in Levchenko (2001), Slepchenko et al. (2002), Kholodenko (2006), Mogilner (2006), and Mogilner et al. (2006).

Modeling Successes

We review recent successful models in order of increasing complexity of the nature of the addressed phenomenon, of the methods used, and of the models' utility. For convenience, the reviewed models are listed and classified in Table 1. The space constraints preclude us from detailed discussions, so a few models (Marshall and Rosenbaum, 2001; Kimura and Onami, 2005; Malikov et al., 2005) are discussed in greater detail in the Supplemental Data (available with this article online) in order to illustrate better how the modeling logic works.

A Simple Mathematical Model Can Be Used as a Quantitative Hypothesis to Be Tested in Future Experiments or Can Simply Be Thought Provoking

The mathematician Georg Polya remarked once that if there is a difficult problem that you cannot solve, then there is also a simpler problem you do not understand. Traditional modeling is an art of abstracting a simpler system, which is a "caricature" rather than a "photograph" of the actual biological system. Complex computer experiments and system-level analyses (see below) are all the rage now, but coarse-grained mathematical models remain useful steps in a reductionist agenda of studying single-scale modules in vast cellular networks when so few details about the actual in vivo processes are known that it is difficult to proceed without numerous (and often arbitrary) assumptions about the nature of nonlinearities and parameter values.

Mathematical modeling is no different from other scientific methodologies. Just like in experimental biology, the important step is to identify an important problem that could benefit from modeling. Data are necessary to make the model grounded in reality, but not sufficient. It is the scientific question motivating the model that is the key for a good model. When a mathematical model is developed, it is often possible to understand it in an intuitive way. That is, one can look at such a model and grasp how the behavior might rely on a certain range of parameters, without actually solving the model equations (scaling and nondimensionalization are invaluable tools: see the Supplemental Data). Such qualitative insight requires very hard and long thinking, so a really good model cannot be cranked up overnight. Thus, biological insight may be an immediate product of building the model, so, like in Zen, it is often the path not the end goal that makes modeling beneficial. The process of modeling by itself forces the modeler to ask many important questions: What are the different parts of the system? Are they all equally important? How do those components interact? So it is the process of writing equations and asking these questions that leads to a new qualitative understanding, which is the real benefit of modeling—after all, we want qualitative understanding, not to replace a biological charade with a mathematical one. The following examples illustrate these philosophical arguments.

Mathematically, the simplest models emerge when one looks for a functional dependence of one observable from another—in that case, algebraic equations (Supplemental Data) suffice. An example of such model is the recent theoretical explanation (Vavylonis et al., 2006) of the surprising observation (Kovar et al., 2006) that in the presense of "leaky capper" formin, the F-actin polymerization rate is an increasing function of small

Table 1. Reviewed Models

Purpose of Model	Biological Process	Temporal Dynamics	Spatial Component	Stochastic	Appropriate Quantitative Technique	Reference
Quantitative hypothesis testing	Actin dynamics	Yes	No	No	Algebraic equations	Vavylonis et al., 2006
	Biochemical oscillator-"clock" in *Myxobacteria*	Yes	No	No	ODEs	Igoshin et al., 2004
	MAPK signaling cascade	Yes	No	No	ODEs	Markevich et al., 2004
	Pattern formation in *E. coli* division	Yes	Yes	No	PDEs	Meinhardt and de Boer, 2001
Data interpretation	Fluorescence recovery after photobleaching (FRAP)	Yes	Yes	No	PDEs	Beaudouin et al., 2006
Data integration	Nucleocytoplasmic transport	Yes	Yes	No	ODEs, PDEs	Smith et al., 2002; Gorlich et al., 2003; Riddick and Macara, 2005
	Regulation of cell cycle in budding yeast	Yes	No	No	ODEs	Chen et al., 2000; Chen et al., 2004
Computer experiment	Cytoskeleton dynamics	Yes	Yes	Yes	Monte-Carlo simulations	Haviv et al., 2006
	Cytoskeleton dynamics	Yes	Yes	Yes	Large-scale agent-based computer simulations	Alberts and Odell, 2004

concentrations of profilin, a decreasing function of large concentrations of profilin, and has a maximum at an intermediate profilin concentration. Profilin facilitates ADP-ATP exchange on actin monomers and shifts the equilibrium to ATP-G-actin-profilin complexes associating with filaments' barbed ends. Vavylonis et al. (2006) hypothesize that both profilin and actin-profilin complexes can bind to one of two formin domains (that remain processively attached at the growing filament's barbed end) and get transferred from formin to the growing tip. It is easy to see that this explanation of an optimal profilin concentration giving a maximal actin elongation rate is very simple: a high profilin concentration suppresses elongation, largely because free profilin displaces profilin-actin from the formin. However, without writing equations and thinking about them, this explanation is not easy to come up with. This model is a novel quantitative hypothesis that can be tested by biochemistry, but it cannot explain some previous observations. Commendably, the authors do not obfuscate this fact but rather discuss, based on the model framework, how these discrepancies can elucidate still murky ATP-hydrolysis mechanisms coupled to actin polymerization.

The next level of mathematical complexity is deterministic temporal dynamics described adequately by ordinary differential equations (ODEs; see Figure 1 and the Supplemental Data)—the most common type of representation used to model cell signaling and metabolic pathways in cases when the cell can be considered as a well-stirred reactor and when stochastic effects can be neglected. Models of this type are especially useful when biological oscillators and switches are considered—only through very advanced intuition coming from years of working with nonlinear dynamical systems can one tell whether a known system of interactions manifests some systems-level property and, if so, how? Two characteristic examples are a model of the C signaling-based reversal biochemical clock of *Myxococcus xanthus* cells (Igoshin et al., 2004) and a model of the mitogen-activated protein kinase (MAPK) cascade bistable switch (Markevich et al., 2004).

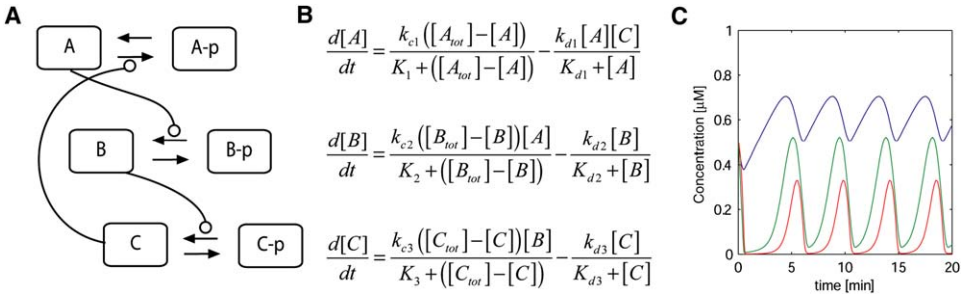

Figure 1. Example of a Simple ODE Model of a Biochemical Network

A simple biochemical network consists of three proteins. Each hypothetical protein can be in one of two states, phosphorylated and unphosphorylated. The network includes three feedback interactions (curves with circles at their ends represent activation of respective reactions) between the different proteins. Even with this simplified network containing three proteins, it is hard to predict based on intuition alone what the network behavior will be. The network layout (A) is implemented as a mathematical model using ODEs with Michaelis-Menten expressions for all reactions (B). Simulations results are shown in (C). A more detailed description of this model, which is a simplified version of the biological clock model of (Igoshin et al., 2004), is presented in the Supplemental Data.

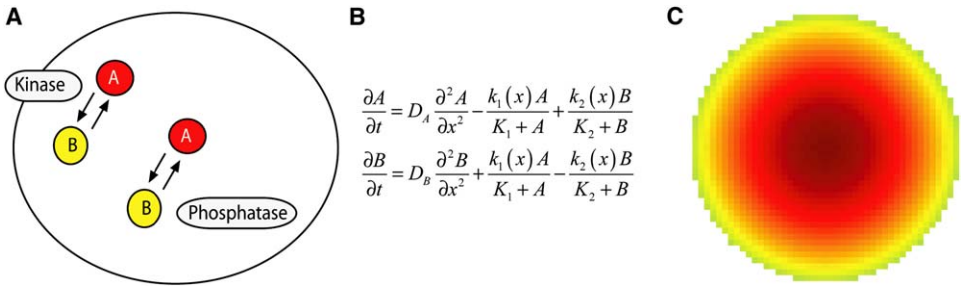

Figure 2. Example of a Simple Spatial Model

In eukaryotic cells, localization of proteins plays an important role in many biological phenomena. In this hypothetical example, a protein can exist in two states, A and B. A kinase that catalyzes the A→B reaction is bound to the cell membrane, and a phosphatase that catalyzes the B→A reaction is localized diffusively throughout the cytoplasm (A). A PDE model of this network (B) includes both diffusion and reaction terms. A hypothetical example of the result of such a model—stable steady spatial distribution of protein in one of the states—is shown in (C). A more detailed description of the model is presented in the Supplemental Data.

Myxobacteria—common soil bacteria that are often studied for their multicellular social behavior—glide back and forth on the surface in the "rippling phase," exhibiting unusual nonrandom reversal time distribution and a "refractory period" after each reversal during which the cell is not responsive to signaling, indicating that the cells have "memory." These observations led Igoshin et al. (2004) to hypothesize an internal biochemical cycle that controls reversals and acts as a clock affected by collisions between cells through C signaling homologous to the chemosensory signal transduction pathway (Che system) of *E. coli*. By pondering the known "interactome" of the C signaling circuit and hypothesizing that part of it constitutes delayed positive feedback, and another part constitutes a negative feedback oscillator, the authors demonstrate that a relatively simple (yet, impossible to intuit without math) kinetic scheme (Figure 1) results in stable oscillations, effectively proposing the quantitative hypothesis about the structure and dynamic of the "clock." Igoshin et al. (2004) use the model to explain all observations, from both wild-type and mutant assays, and propose experiments to test the model.

Bistability is a ubiquitous principle of cellular regulation when pathways display switch-like behavior in response to a stimulus. Positive or double-negative feedback regulation is generally considered to be a prerequisite for bistability, so at first glance, bistability cannot arise at, say, an individual kinase level unless there is allosteric activation or inhibition of a converter enzyme by its product or substrate, creating a positive circuit required for bistability. By mathematically investigating MAPK pathways, which are critical for cellular decisions to proliferate, differentiate, or undergo apoptosis, Markevich et al. (2004) discovered that a dual phosphorylation-dephosphorylation cycle for the kinase and phosphatase already possesses all the required ingredients to display bistable behavior. This conclusion is invaluable as a reference point for future experimental work on cell signaling pathways.

Dynamical systems models are useful, not only when the biological dynamics are complex, but even when the system is exceedingly simple mathematically, for instance, when a simple stable steady state is observed. One such mathematical model (Marshall and Rosenbaum, 2001) provides clues about the molecular activities that control flagellar length and is reviewed in detail in the Supplemental Data.

A unique feature of cells is the utilization of spatial separation of components and interactions. Partial differential equations (PDEs, which are much more complex than ODEs; see the Supplemental Data and Figure 2) can be used to model spatially heterogeneous and compartmentalized dynamics. An excellent example of such model is the theory of MinC/MinD spatial-temporal oscillations (Meinhardt and de Boer, 2001). In bacteria, the division plane is determined by a polymeric ring of the FtsZ protein. The site of this ring assembly is controlled by the Min system, suppressing FtsZ polymerization away from the center. It was observed that the Min proteins in *E. coli* undergo unusual oscillations between the membrane of both cell halves. Meinhardt and de Boer (2001) showed that they can reproduce the oscillatory pattern if the following assumptions are made: (1) the MinD ATPase self-assembles on the membrane and recruits both MinC, an inhibitor of Z ring formation, and MinE, a protein required for MinC/MinD oscillation; (2) a local accumulation of MinE is generated by a reaction based on local self-enhancement and a long-range antagonistic effect; and (3) MinE displaces MinD from the membrane, causing its own local destabilization and shift toward higher MinD concentrations. By translating these assumptions into a system of PDEs and solving them numerically, the authors demonstrated that this local destabilization results in a wave of high MinE concentration traveling from the cell center to a pole, where it disappears. MinD reassembles on the membrane of the other cell half and attracts a new accumulation of MinE, causing a wave-like disassembly of MinD, again resulting in a pole-to-pole oscillation of MinC/MinD. On the average, MinC concentration is highest at the poles, forcing FtsZ assembly to the center. Yet again, after the model is suggested, it can easily be understood without math, but not before the equations are written and investigated. This model's value is in explaining the system as a self-organizing one, not requiring any prelocalized "morphogen gradients," yet maybe the greatest achievement of the Meinhardt and de Boer's paper is that it started (together with Howard et al., 2001) an avalanche of increasingly sophisticated

and realistic models, addressing, among others, questions about the role of stochastic effects in pattern formation (reviewed in Howard and Kruse, 2005).

A Model Can Be a Tool for Data Interpretation

Often, it is hard to interpret raw data. Quantitative models can be useful in analyzing complex measurements and extracting rates. For example, *f*luorescent *r*ecovery *a*fter *p*hotobleaching (FRAP) is often the method of choice for analyzing the kinetics of binding of fluorescently labeled proteins. However, the analysis of FRAP measurements is not trivial, and many FRAP studies use simplifications by ignoring boundaries and assuming infinite volume and neglecting diffusion, assuming that it is much slower than the reaction studied. Those simplifications are not always justified, often, the bleach region is comparable in size to the volume in which molecules can diffuse, and many reactions are as fast as diffusion.

Using explicit description of both the diffusion and reaction components and solving respective PDEs, Beaudouin et al. (2006) overcame the need to simplify. The mathematical model used is very simple conceptually: the protein is free to diffuse within the nucleus with some characteristic diffusion coefficient and can bind and unbind from DNA with appropriate kinetic parameters. Using an explicit mathematical model rather than simplifying assumptions, the authors showed that very fast kinetics cannot be distinguished from diffusion, suggesting that DNA binding protein might perform a search which can be approximated by effective one-dimensional diffusion along the DNA filament, rather then a three-dimensional diffusion in the entire nuclear volume. Another beautiful example of using modeling to interpret data is recent investigation of the leading edge dynamics in motile cells by Waterman-Storer's and Danuser's labs (Ponti et al., 2003).

Models as Tools for Data Integration and Understanding

Experimental studies increasingly expose complex cellular networks with potentially vast arrays of rates characterizing them. Which of these rates are the most important to measure accurately? A good example of such a system is the transport in and out of the nucleus (Smith et al., 2002; Gorlich et al., 2003; Riddick and Macara, 2005), the chemical environment of which is controlled by the nuclear pore complexes (NPC) and an extensive protein system. To allow such transport, the cell uses the GTPase Ran. The Ran guanine-nucleotide exchange factor (RanGEF), Rcc1, is localized in the nucleus and converts RanGDP into RanGTP, while the Ran GTPase-activating protein (RanGAP) that catalyzes the RanGTP to RanGDP transition localizes to the cytoplasm. These differences in localization of the "source" and "sink" create a concentration gradient of RanGTP across the nuclear membrane (Figure 2). This gradient is utilized to create directionality of both import and export using an array of importin and exportin proteins to determine specificity of transported molecules (Pemberton and Paschal, 2005). Nucleocytoplasmic transport is very complex: there are more than 15 import/export carriers and several additional accessory proteins that create and maintain the RanGTP gradient across the nuclear membrane. To gain an insight, a few groups have used quantitative modeling to explore nucleocytoplasmic transport at a system level, as a system of coupled integrated modules.

The initial modeling attempt (Smith et al., 2002) used several simplifications. Only a few important aspects of the transport machinery were modeled; for example, all the different import and export carriers were represented by a single generic "carrier" protein. In addition, multistep reactions were simplified into a single step assuming Michaelis-Menten kinetics. Even this simplified model provided insights into the relative importance of different parts of the system. For example, sensitivity analysis (testing how much the results vary when model parameters are changed systematically) revealed that the transport rate was most sensitive to changes in RanGEF's catalysis of the RanGDP to RanGTP exchange. The second attempt (Gorlich et al., 2003) used similar tools (ODEs derived from wiring diagram of the changes in the localization and states of molecules) but different simplifications and reached a different conclusion: that changes in RanGEF are less important than the permeability of the membrane to RanGDP. Gorlich et al. (2003) explicitly modeled all reaction steps instead of the Michaelis-Menten simplification; however, they did not include the carriers but just calculated the "potential" in the form of the resulting RanGTP gradient. Careful comparison between the two models shows that the difference in the results is not a consequence of the different simplifications but rather a change in the numerical value of one of the parameters in the models: the permeability of the membrane to RanGDP. Without modeling, it would be practically impossible to determine which factor limits transport. The difference between the models suggested which of the estimated parameters in this system is most significant and should be measured with greater experimental accuracy. Additional experimental work with a next generation, more detailed model (Riddick and Macara, 2005) resolved this issue and demonstrated that the RanGEF is unlikely to be the limiting factor and that the permeability and concentration of Ran and other accessory molecules are more likely candidates. The study of the Ran transport system is a good example of embracing cellular complexity, in which modeling assisted experiments that lead to a second generation of more detailed models.

Increasingly Complex Generations of Models Can Be Used to Understand Cellular Networks as Systems

Another spectacular modeling achievement is the recent development of mathematical models of budding yeast cell cycle (Chen et al., 2004), one of the first true successes of system biology. This model made the transition from reductionism to synthesis by integrating simple signaling modules into multiscale pathways. The budding yeast cell cycle was extensively studied, and the respective control system is known in exquisite detail in wild-type cells and more than a hundred mutants, so intuition cannot be used to understand the system behavior anymore. Chen et al. (2004) integrated data from tens of independent studies into a qualitative consensus model of the cell cycle expressed in the form of a "wiring" network diagram, and they translated the diagram into a set of nonlinear ODEs and tested the validity of the mechanistic model by solving the equations numerically. One of the most amazing features of the

cell cycle model was that so few quantitative parameters were known that its authors had to use genetics to guess a set of reasonable, internally consistent parameters to use in the model. The solutions were compared with experimental results from 131 different yeast mutant strains. In its current form, the model can explain the results of 120 mutant strains, which is a definitive success, suggesting that Chen et al. have identified all essential players and interactions. The model predicts phenotypes of new mutant combinations and estimates many reaction rates that are difficult to measure. This model also shows which of the complex network modules, types of reactions, and reaction rates are critically important and which are not, thereby making important proposals for future experimental work. The power of modeling, however, is often not what you can explain but rather what you cannot—the unexplained 11 mutants indicate specifically which aspects of the mechanism require revisions. Note, that the model's inability to account for 11 mutants may be due to defects, not in the model, but in the experimental design and/or interpretation.

It is important to realize that the success of the budding yeast cell cycle model would be impossible without earlier, simpler models from the same group; Tyson et al. (1995) suggested an almost primitive model of two equations that illustrated important general principles, and then Chen et al. (2000) suggested a much more elaborate model, albeit with a primitive "exit from mitosis" module. These models helped to build intuition and develop quantitative modules of the whole network without which the complete model (Chen et al., 2004) would be incomprehensible.

Characteristically, the early simple models of the cell cycle were most popular with mathematical biologists but were completely ignored by molecular cell biologists, who could not take them seriously because so many known components of the control mechanism were missing. This dichotomy is one of the barriers to modeling in biology, and perhaps the only solution is to develop both simple and system-level models and to work back and forth across different modeling scales. The alternative approach—to build the system-level models right away by assembling a huge network diagram from the data scattered in tens of papers, then translating the diagram into tremendous sets of equations, and then using such models to look for undiscovered dynamics—has dubious value. One recent example of such an approach is the model of TLR and IL-1R signaling networks responsible for immune response (Oda and Kitano, 2006). The model consists of 652 variables describing molecular species with 444 reactions between them and an astronomical number of parameters. It is hard to comprehend how such model can be used and how the reliability of its conclusions can be assessed.

From the above, it should be clear that the translation of a network diagram into a set of mathematical equations is a crucial step, because there are multiple ways to do that. There are two types of decisions that need to be made: first, what type of reaction do each of the arrows represent (e.g., mass action, Michaelis-Menten, etc.), and second, what are the specific kinetic parameter values characterizing the reaction. With tens of equations and kinetic parameters in modern models, the gap between the cartoon and mathematical representations is tremendous. Fortunately, the large number of estimated parameters does not necessarily mean that the model is flexible enough to generate any type of result. The specific form of the equations themselves restricts the possible result. Still, in the future, different parameter combinations will have to be tested systematically in order for the system-level models to become truly predictive. Very recently, a number of groups started to make promising progress in this direction by using genetic and stochastic optimization algorithms to automatically explore the whole model parameter space (Mezer et al., 2006). Such efforts require a significant amount of quantitative data to be used to "score" the model's success, so modeling efforts become inseparable from experimental ones.

Computer Experiments Can Confirm the Plausibility of a Qualitative Model or Explore a Complex Phenomenon When There Is Little Intuition about It

Sir Isaac Newton once said, "Equations are smarter than us"—meaning that formal mathematical solutions suggest qualitative answers that cannot be foreseen without math. This suggests an interesting approach to modeling: not to theorize much but rather to build a computer model by formally translating qualitative assumptions into computational rules, and see how it behaves, in the same way that a chess programmer writes a program and then watches it play. This approach is fruitful if the modeled phenomenon, on the one hand, yields qualitative understanding, but on the other hand, involves a multidimensional process taking place in time and space. If, in addition, the discrete nature of the molecular components cannot be neglected, then traditional approximations and analyses of ODEs or PDEs are too difficult and often fruitless. Such a model—a computer experiment, really—is especially useful for studying interrelationships between microscopic and macroscopic behaviors in a manner that is difficult or impossible to do in the lab. This could require a lot of computational power, but given the constantly increasing speed and decreasing cost of computers, this is not a limitation anymore.

One example of such a study is kinetic Monte Carlo (see the Supplemental Data and Figure 3) simulation of actin aster-like structures (Haviv et al., 2006). This study addresses a question about how a motile cell "chooses" an actin network type: some cells emphasize largely isotropic lamellipodial networks in which actin filaments branch off of each other at wide angles, whereas other cells are dominated by filopodia—highly aligned actin filament bundles. The authors investigated in vitro self-assembly of actin-based structures and discovered that Arp2/3 complex and fascin control the network type; without fascin, the Arp2/3 complex mediates branching of nascent "daughter" filaments from existent "mother" filaments, resulting in spontaneous formation of diffuse aster-like structures. In the presence of the bundling protein fascin, these asters transition into stars with filopodia-like bundles of actin filaments growing from the surface.

The authors came up with an elegant idea that the microscopic mechanism of actin reorganization is fascin-mediated bending and bundling of longer filaments

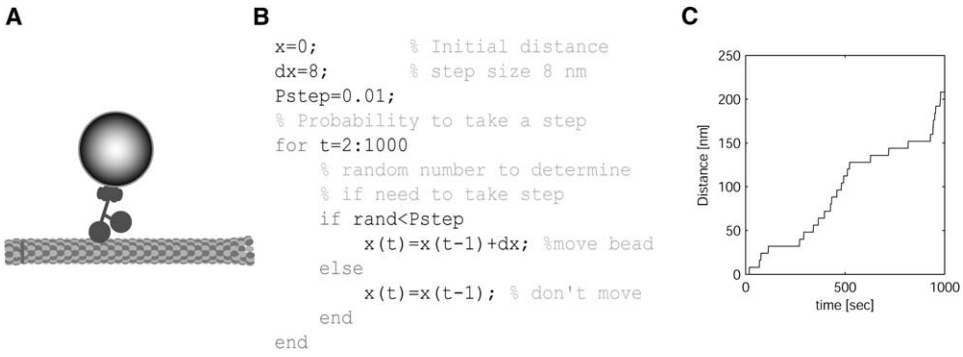

Figure 3. Example of a Simple Stochastic Monte Carlo Model

(A) An example of a molecular motor "walking" along a microtubule with a vesicle as cargo. This process can be easily modeled using stochastic computer simulations.

(B and C) Computer code (B) can simulate a single walk of the motor and generate a trajectory of distance over time (C). A more detailed description of the model is presented in the Supplemental Data.

and coalescence of other long filaments into these nascent bundles. Microscopy resolution is insufficient to test this hypothesis, so a model was used for in silico experiments, testing it and complementing the in vitro part of the study. The simulation, indeed, revealed that in the in silico system of filaments that branch off the sides of existent filaments polymerize, get capped, bend, and get crosslinked by fascin with known rates (luckily, the kinetics and mechanics of actin are studied thoroughly), the emerging patterns resemble the observed ones. As a bonus, modeling makes an intuitively clear prediction that capping protein inhibits star formation—shorter filaments cannot bend and bundle—which is observed, and a much less intuitive prediction that the transition from asters to stars depends on the ratio of fascin to total amount of actin, which is also observed.

Technically very similar simulations were used to decide which of the two mechanisms—"pushing" or "pulling"—is responsible for nuclear centering in *C. elegans* (Kimura and Onami, 2005). Simulations predicted that the kinetics of the centering process are very different for these two mechanisms. Remarkably, the experiment showed that the data support the pulling mechanism's prediction. In this case, modeling helped to choose between two equally plausible hypotheses, and without it the kinetic data are almost useless (see the detailed discussion in the Supplemental Data).

In the previous example, behavior is not too complex, and in principle the theoretical predictions could be made without elaborate simulations (but probably not without any modeling). Computer experiments are even more useful when behavior is so complex that intuition fails. In such cases it is sometime a good idea to simulate an agent-based model, in which each molecule is represented explicitly and interactions between them obey simple rules of physics and chemistry. The complexity comes from the great heterogeneous number of interacting players. The beauty of this approach is that it requires little pondering (but often a lot of programming). The problem is that such an exceedingly complicated, fine-grained model, even if it can recapitulate a known cell biological phenomenon, may not by itself yield any insight into the principles of the mechanism. Large-scale order may be seen to arise within a computer model, but understanding how it arises may be no easier than understanding how it arises in the living cell in the first place. Yet, the key difference between the complex simulation and the complex biological system is that the model's designer has exquisite control over every aspect of the system, something that any experimental biologist would trade an arm for. So, the undeniable utility of such a model is that by playing with interaction rules and comparing results to the data, one can find what the essential molecular mechanisms are even without detailed understanding of how they work.

One of the best examples of such an approach is the first in silico reconstruction of *Listeria*'s movement (Alberts and Odell, 2004). *Listeria* moves by hijacking the host cell's actin system to grow a comet-like tail of actin filaments that pushes the bacterium through the cytoplasm (Cameron et al., 2000). In this model, actin filaments both propel the virtual bacterium by polymerizing against its surface and pushing it and resist its movement by attaching to the surface transiently. Each actin monomer and actin accessory protein is simulated explicitly according to relatively well-known rules of actin dynamics. By accounting for realistic geometry, actin network architecture, and stochastic processes, which earlier, more coarse-grained models could not address (Mogilner, 2006), the authors discovered that *Listeria* moves in irregular nano-steps—the phenomenon actively investigated experimentally—even though the nature of these steps is still murky. The simulations also result in a vivid and realistic mimicking of *Listeria*'s propulsion.

The Closer the Complete Understanding of Cell Behavior Is, the Closer Experiment and Theory Become, and the Question about Utility of Modeling Fades Away

Probably the most tractable phenomenon for systems biology is *E. coli* chemotaxis (Baker et al., 2006). This relatively simple system is built from a small number of proteins with quantitative structural and biochemical details now available for every node and link in the wiring diagram of the sensory pathway. These vast arrays of physiological and biochemical data allow mathematical models of chemotaxis and signal kinetics to be

generated. Already 10 years ago these models were so detailed (Spiro et al., 1997; Barkai and Leibler, 1997), with very little wiggle room to make arbitrary assumptions or parameter choices, that quantitative experimental analysis of bacterial chemotaxis has benefited from theory ever since. Among other benefits, the concept of robustness—the ability of a signaling system to tolerate variations in protein concentrations and reaction rates—was understood quantitatively and applied to many other signaling pathways in prokaryotes and eukaryotes. Recently, truly system-level models (Shimizu et al., 2003; Rao et al., 2004) almost reached the goal of becoming the "end point" of biological research (these models account for most observations in chemotaxis, but they still have to be "fine tuned" to match the data more closely). In fact, the bacterial chemotaxis field has gotten past modeling single cells to consider the behavior of populations of cells. For example, Korobkova et al. (2004) combined experimental measurements and computer simulations to demonstrate that some of the parameters of the sensory system are set in such a way that the bacterial population responds to chemical signals stochastically and nonuniformly, and that such a response has survival value. Characteristically, researchers in this field do not even discuss anymore the experiment-theory relationship; those merged recently without any hype, much as theory merged with experiment in physics more than a century ago. A few other cell biological fields, e.g., cell migration, are not far behind, and there are other areas of biology where the merger took place long ago, such as biochemical kinetics, neurophysiology, and ion channel biophysics.

The Future of Quantitative Modeling in Cell Biology

There are significant barriers to modeling in cell biology. A lack of standard modeling tools needed to reproduce modeling results (Cassman, 2005) is not one of them: reproducibility is not a big problem for a relatively simple model, for which it is enough to grasp its essentials and then utilize it in a somewhat changed form. A number of groups have already developed standard interfaces enabling biologists to build large-scale computational models, run simulations, and visualize simulation results in a way that allows direct comparison to experiments (Supplemental Data). One of the obstacles is that experimental biologists are not typically equipped with computational expertise sophisticated enough to generate quantitative predictions from models. On the other hand, modelers often have difficulty in deciding which experiments are reliable (the experimental literature in biology is a minefield of failures in experimental design and interpretation), and they often lack understanding of experimental techniques and, more importantly, of what are important biological questions. The worst problem is mutual suspicion: many experimentalists still think that modeling is but "window dressing" and that modelers "parasitize" hard-earned data, while many modelers cannot get used to the fast pace of biological research and complain that their work is not cited enough. All these difficulties can and will be overcome; curricula are being developed that have already started to produce a new generation of researchers equally comfortable with computation and bench work and who have double biological/computational mentality. Most importantly, a number of "dry" and "wet" labs are engaged in a "closed loop," in which experiments inspire models, the predictions of which lead to further experiments, causing refining of the next generation models.

The role of quantitative modeling in biology as a complementary instrument of biological discovery will continue to increase, but in our opinion it will always be a peculiar tool, not equal to biochemistry, microscopy, and genetics—an "art" as much as science. Building a good model depends too much on intuition, on the rare abilities to ask the right question and to sense mathematical order behind messy facts, on tricky timing (not too early, when absence of data leaves a model unsubstantiated, not too late, when everything is clear without a model), and on hard, long thinking that makes modeling so painful, but also so much fun.

Supplemental Data

Supplemental Data include mathematical background and resources and supplemental examples and can be found with this article online at http://www.developmentalcell.com/cgi/content/full/11/3/279/DC1/.

Acknowledgments

This work was funded by grants from the NIH (NIGMS GM068952-01 and NIGMS U54 GM64346) and the NSF (DMS 0315782) to A.M., a grant from the NSF (MCB 0416310) to W.F.M., and a University of California Systemwide Biotechnology Research & Education Program GREAT Training Grant (2006-03) to R.W. We are grateful to anonymous reviewers for insightful suggestions.

References

Alberts, J.B., and Odell, G.M. (2004). In silico reconstitution of Listeria propulsion exhibits nano-saltation. PLoS Biol. *2*, e412.

Baker, M.D., Wolanin, P.M., and Stock, J.B. (2006). Systems biology of bacterial chemotaxis. Curr. Opin. Microbiol. *9*, 187–192.

Barkai, N., and Leibler, S. (1997). Robustness in simple biochemical networks. Nature *387*, 913–917.

Beaudouin, J., Mora-Bermudez, F., Klee, T., Daigle, N., and Ellenberg, J. (2006). Dissecting the contribution of diffusion and interactions to the mobility of nuclear proteins. Biophys. J. *90*, 1878–1894.

Cameron, L.A., Giardini, P.A., Soo, F.S., and Theriot, J.A. (2000). Secrets of actin-based motility revealed by a bacterial pathogen. Nat. Rev. Mol. Cell Biol. *1*, 110–119.

Cassman, M. (2005). Barriers to progress in systems biology. Nature *438*, 1079.

Chen, K.C., Csikasz-Nagy, A., Gyorffy, B., Val, J., Novak, B., and Tyson, J.J. (2000). Kinetic analysis of a molecular model of the budding yeast cell cycle. Mol. Biol. Cell *11*, 369–391.

Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., and Tyson, J.J. (2004). Integrative analysis of cell cycle control in budding yeast. Mol. Biol. Cell *15*, 3841–3862.

Field, R.J., and Noyes, R.M. (1974). Oscillations in chemical systems, part 4. Limit cycle behavior in a model of a real chemical reaction. J. Chem. Phys. *60*, 1877–1944.

Gliksman, N.R., Skibbens, R.V., and Salmon, E.D. (1993). How the transition frequencies of microtubule dynamic instability (nucleation, catastrophe, and rescue) regulate microtubule dynamics in interphase and mitosis: analysis using a Monte Carlo computer simulation. Mol. Biol. Cell *4*, 1035–1050.

Gorlich, D., Seewald, M.J., and Ribbeck, K. (2003). Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation. EMBO J. *22*, 1088–1100.

Haviv, L., Brill-Karniely, Y., Mahaffy, R., Backouche, F., Ben-Shaul, A., Pollard, T.D., and Bernheim-Groswasser, A. (2006). Reconstitution of the transition from lamellipodium to filopodium in a membrane-free system. Proc. Natl. Acad. Sci. USA *103*, 4906–4911.

Howard, M., and Kruse, K. (2005). Cellular organization by self-organization: mechanisms and models for Min protein dynamics. J. Cell Biol. *168*, 533–536.

Howard, M., Rutenberg, A.D., and de Vet, S. (2001). Dynamic compartmentalization of bacteria: accurate division in E. coli. Phys. Rev. Lett. *87*, 278102.

Igoshin, O., Goldbeter, A., Kaiser, D., and Oster, G. (2004). A biochemical oscillator explains several aspects of M. xanthus behavior during development. Proc. Natl. Acad. Sci. USA *101*, 15760–15765.

Janes, K.A., Gaudet, S., Albeck, J.G., Nielsen, U.B., Lauffenburger, D.A., and Sorger, P.K. (2006). The response of human epithelial cells to TNF involves an inducible autocrine cascade. Cell. *124*, 1225–1239.

Janetopoulos, C., Ma, L., Devreotes, P.N., and Iglesias, P.A. (2004). Chemoattractant-induced phosphatidylinositol 3,4,5-trisphosphate accumulation is spatially amplified and adapts, independent of the actin cytoskeleton. Proc. Natl. Acad. Sci. USA *101*, 8951–8956.

Joyce, A.R., and Palsson, B.O. (2006). The model organism as a system: integrating 'omics' data sets. Nat. Rev. Mol. Cell Biol. *7*, 198–210.

Kholodenko, B.N. (2006). Cell-signalling dynamics in time and space. Nat. Rev. Mol. Cell Biol. *7*, 165–176.

Kimura, A., and Onami, S. (2005). Computer simulations and image processing reveal length-dependent pulling force as the primary mechanism for *C. elegans* male pronuclear migration. Dev. Cell *8*, 765–775.

Korobkova, E., Emonet, T., Vilar, J.M., Shimizu, T.S., and Cluzel, P. (2004). From molecular noise to behavioural variability in a single bacterium. Nature *428*, 574–578.

Kovar, D.R., Harris, E.S., Mahaffy, R., Higgs, H.N., and Pollard, T.D. (2006). Control of the assembly of ATP- and ADP-actin by formins and profiling. Cell *124*, 423–435.

Levchenko, A. (2001). Computational cell biology in the post-genomic era. Mol. Biol. Rep. *28*, 83–89.

Ma'ayan, A., Blitzer, R.D., and Iyengar, R. (2005). Toward predictive models of mammalian cells. Annu. Rev. Biophys. Biomol. Struct. *34*, 319–349.

Malikov, V., Cytrynbaum, E.N., Kashina, A., Mogilner, A., and Rodionov, V. (2005). Centering of a radial microtubule array by translocation along microtubules spontaneously nucleated in the cytoplasm. Nat. Cell Biol. *7*, 1213–1218.

Markevich, N.I., Hoek, J.B., and Kholodenko, B.N. (2004). Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. J. Cell Biol. *164*, 353–359.

Marshall, W.F., and Rosenbaum, J.L. (2001). Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. J. Cell Biol. *155*, 405–414.

Meinhardt, H., and de Boer, P.A. (2001). Pattern formation in Escherichia coli: a model for the pole-to-pole oscillations of Min proteins and the localization of the division site. Proc. Natl. Acad. Sci. USA *98*, 14202–14207.

Mezer, A., Ashery, U., Gutman, M., Project, E., Bosis, E., Fibich, G., and Nachliel, E. (2006). Systematic search for the rate constants that control the exocytotic process from chromaffin cells by a genetic algorithm. Biochim. Biophys. Acta *1763*, 345–355.

Mogilner, A. (2006). On the edge: modeling protrusion. Curr. Opin. Cell Biol. *18*, 32–39.

Mogilner, A., Wollman, R., Civelekoglu-Scholey, G., and Scholey, J. (2006). Modeling mitosis. Trends Cell Biol. *16*, 88–96.

Oda, K., and Kitano, H. (2006). A comprehensive map of the toll-like receptor signaling network. Mol. Syst. Biol. *2*, E1–E20.

Pemberton, L.F., and Paschal, B.M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic *6*, 187–198.

Ponti, A., Vallotton, P., Salmon, W.C., Waterman-Storer, C.M., and Danuser, G. (2003). Computational analysis of F-actin turnover in cortical actin meshworks using fluorescent speckle microscopy. Biophys. J. *84*, 3336–3352.

Prill, R.J., Iglesias, P.A., and Levchenko, A. (2005). Dynamic properties of network motifs contribute to biological network organization. PLoS Biol. *3*, e343.

Rao, C., Kirby, J., and Arkin, A. (2004). Design and diversity in bacterial chemotaxis: a comparative study in Escherichia coli and Bacillus subtilis. PLoS Biol. *2*, E49.

Riddick, G., and Macara, I.G. (2005). A systems analysis of importin-{alpha}-{beta} mediated nuclear protein import. J. Cell Biol. *168*, 1027–1038.

Schneider, I.C., and Haugh, J.M. (2005). Quantitative elucidation of a distinct spatial gradient-sensing mechanism in fibroblasts. J. Cell Biol. *171*, 883–892.

Segel, L.A. (1988). On the validity of the steady state assumptions of enzyme kinetics. Bull. Math. Biol. *50*, 579–593.

Shimizu, T.S., Aksenov, S.V., and Bray, D. (2003). A spatially extended stochastic model of the bacterial chemotaxis signalling pathway. J. Mol. Biol. *329*, 291–309.

Slepchenko, B.M., Schaff, J.C., Carson, J.H., and Loew, L.M. (2002). Computational cell biology: spatiotemporal simulation of cellular events. Annu. Rev. Biophys. Biomol. Struct. *31*, 423–441.

Smith, A.E., Slepchenko, B.M., Schaff, J.C., Loew, L.M., and Macara, I.G. (2002). Systems analysis of Ran transport. Science *295*, 488–491.

Spiro, P.A., Parkinson, J.S., and Othmer, H.G. (1997). A model of excitation and adaptation in bacterial chemotaxis. Proc. Natl. Acad. Sci. USA *94*, 7263–7268.

Tyson, J.J. (1976). The Belousov-Zhabotinskii Reaction. Lecture Notes in Biomathematics, Volume 10 (Berlin: Springer-Verlag).

Tyson, J.J., Novak, B., Chen, K., and Val, J. (1995). Checkpoints in the cell cycle from a modeler's perspective. Prog. Cell Cycle Res. *1*, 1–8.

VanBuren, V., Cassimeris, L., and Odde, D.J. (2005). Mechanochemical model of microtubule structure and self-assembly kinetics. Biophys. J. *89*, 2911–2926.

Vavylonis, D., Kovar, D.R., O'Shaughnessy, B., and Pollard, T.D. (2006). Model of formin-associated actin filament elongation. Mol. Cell *21*, 455–466.

Weinreb, G.E., Elston, T.C., and Jacobson, K. (2006). Causal mapping as a tool to mechanistically interpret phenomena in cell motility: application to cortical oscillations in spreading cells. Cell Motil. Cytoskeleton, in press. Published online June 23, 2006. 10.1002/cm.20143.

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., and Margalit, H. (2004). Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc. Natl. Acad. Sci. USA *101*, 5934–5939.

Developmental Cell *11*

Supplemental Data

Quantitative Modeling in Cell Biology:

What Is It Good for?

Alex Mogilner, Roy Wollman, and Wallace F. Marshall

Mathematical Background and Resources

1. Algebraic Equations

Algebraic equations, as well as other basic mathematical tools that every biologist has to be comfortable with can be learned almost painlessly from a number of good elementary textbooks, for example (Adler, 2005). A good example of a mathematical model that needs mere algebraic equations is a very simple model of the so called equilibrium single-stranded polymer (Howard, 2001). The assumption of this model is that all the monomer-addition reactions have the same dissociation equilibrium constant, K . That is,

$$A_n + A_1 \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} A_{n+1} \qquad K = \frac{k_{off}}{k_{on}} \tag{1}$$

where A_1 and A_n denote the monomer and n-mer, respectively. In the equilibrium, the concentrations of the polymers are constrained by the infinite system of nonlinear algebraic equations that stem from the balance of fluxes depicted in (1):

$$\frac{[A_n]\times[A_1]}{[A_{n+1}]} = \frac{k_{off}}{k_{on}} = K, n = 1, 2, ... \tag{2}$$

In general, a nonlinear infinite system (2) is very difficult to solve, but in this particular case it is very easy to *guess* the solution (mathematicians call it *ansatz*):

$$[A_n] = Ka^n, n = 1, 2, ... \tag{3}$$

where constant a has to be found from the condition of conservation of the total concentration of subunits:

$$[A_{tot}] = \sum_{n=1}^{\infty} n[A_n] = K\sum_{n=1}^{\infty} na^n = \frac{Ka}{(1-a)^2} = \text{ const} \tag{4}$$

For $\frac{[A_{tot}]}{K} >> 1$, the solution of this quadratic equation: $a \approx 1 - \sqrt{K/[A_{tot}]}$ allows to estimate the average polymer length (average number of subunits in n-mers longer than monomer):

$$n_{av} = \sum_{n=2}^{\infty} n[A_n] / \sum_{n=2}^{\infty} [A_n] = 1 + \frac{1}{1-a} \approx \sqrt{\frac{[A_{tot}]}{K}} \tag{5}$$

2. Ordinary Differential Equations

Excellent introduction into ODEs and their applications in biology can be found in (Edelstein-Keshet, 1988; Lin and Segel, 1974; Logan, 1977).

In Fig. 1 in the main text we show the kinetic scheme of the assumed reactions between proteins FrzE, FrzCD and FrzF in the "Frzlator" model (Igoshin et al., 2004). (We simplified the model considerably by omitting external signals and changing all protein modifications to be phosphorylation.) First, the model variables – concentrations $[A],[B],[C]$ of FrzE, FrzCD and FrzF, respectively, are introduced. Then, using the condition of conservation of the total number of proteins, the concentrations of the phosphorylated proteins are expressed as differences between total conserved concentrations and respective non- phosphorylated concentrations. Then, each arrow in the model diagram is assigned a Michaelis-Menten reaction term in the right-hand side of the respective equation. For example, the rate of change of the concentration of the non-phosphorylated FrzE ($d[A]/dt$) is equal to the phosphorylation rate (proportional to Michaelis-Menten expression $([A_{tot}]-[A])/(K_1+([A_{tot}]-[A]))$) minus de-phosphorylation rate (proportional to Michaelis-Menten expression $[A]/(K_{d1}+[A])$ with the proportionality coefficient $k_{d1}[C]$ reflecting the assumption that this reaction is facilitated by FrzF). Similar arguments lead to three equations for three unknown variables. These nonlinear equations cannot be solved analytically, but qualitative analysis (phase plane analysis, bifurcation theory, perturbation theory, asymptotic analysis) allow finding regions in parameter space where certain behavior, i.e. limit cycle oscillations, can be expected. Ultimately, the system of ODEs has to be solved numerically, which is an almost trivial task these days given availability of huge arrays of computer ODE solvers (see below). Such solutions (see Fig. 1) can be compared with experimental time series for protein concentrations.

3. Partial Differential Equations

Excellent introduction into PDEs and their applications in biology can be found in (Edelstein-Keshet, 1988; Lin and Segel, 1974; Logan, 1977). In Fig. 2 in the main text we illustrate widely accepted model of 'morphogen gradient' which the cell can maintain if kinase and phoshatase are separated spatially (Kholodenko, 2006). (Again, we simplified the models reported in the literature.) Let $B(x,t)$ be the concentration of the phosphorylated (active) protein (for simplicity, we consider a one-dimensional spatial case), and $A(x,t)$ be the concentration of the de-phosphorylated (inactive) protein. The left hand sides of the PDEs shown in Fig. 2 are the rates of change of the concentrations at specified spatial locations. The first terms in the right hand sides describe diffusion of the protein species in the cytoplasm. The second terms are responsible for chemical kinetics of phosphorylation and de-phosphorylation reactions, similar to those described in the ODE example. Coefficients $k_{1,2}(x)$ account for spatial distributions of kinase and

phosphatase, respectively. These two equations have to be complemented with boundary conditions, for example no-flux conditions at the cell boundary. The system of these two equations can be solved numerically without a problem with the help of readily available software (see below). Also, if one is only interested in the steady state solution, PDEs can be converted into ODEs: when transient temporal changes die out, the left-hand sides of the PDEs become zeros, and partial derivatives at the right-hand sides become ordinary ones. Then, ODEs can be solved numerically. The solutions (like the one shown in the figure) can be compared to experimental fluorescent signal, for example.

Note that sometime more complex behavior than just transient relaxation to a single spatial pattern can be expected. For example, for certain values of model parameters, spatially uniform distribution of protein is the only stable state, while in other regions of the parameter space the uniform distribution looses its stability, and temporally steady, spatially periodic stable distributions evolve (this is the case of famous Turing pattern formation, ubiquitous in biology). Even more complex, periodic both in time and space, patterns, can emerge, like in the case reviewed in the main text. In this situation, besides numerical analysis, qualitative analysis (linear stability analysis and the ODE-related tools mentioned) is very useful.

4. Scaling and Nondimensionalization

Excellent introductions into scaling and non-dimensionalization can be found in (Lin and Segel, 1974; Logan, 1977). These tools are largely 'art' rather than algorithmic mathematical methods, and can be only understood by doing many exercises. The exertion is worth it: scaling and non-dimensionalization often allows significant insight without actually solving equations. For example, in the case of algebraic equations described above, it is easy to realize that the dissociation equilibrium constant, K, is the natural scale of concentrations. Introducing the dimensionless concentrations $a_n = [A_n]/K$, (2) can be re-written in simpler form: $\frac{a_n \times a_1}{a_{n+1}} = 1, n = 1, 2, \ldots$ and it has simpler solution. Moreover, an experienced applied mathematician would immediately realize at this point that there are two qualitatively different regimes in this system: one, when the total subunit concentration is much smaller than K, and another, when it is much greater than K. In the former case, almost all subunits are in monomeric form, while in the latter case they are polymerized. These important conclusions do not require actual solutions.

In the ODE model case, one just has to realize that $[A_{tot}], [B_{tot}], [C_{tot}]$ are the natural scales for the respective concentrations, while inverted rate constants ${k_{d2}}^{-1}$ or ${k_{d3}}^{-1}$ are the natural temporal scales. Using these scales to non-dimensionalize the time and concentrations in the equations, first, decreases the number of parameters in the model from 15 to 11 making it easier to explore the parameter space. More importantly, a few non-dimensional parameter combinations evolve, some of which are bound to be either much

smaller or much greater than unity, which normally hints at biologically important hierarchy of time scales in the system.

In the PDE model, assuming that $k_1(x) \sim k_2(x) \sim k, K_1 \sim K_2 \sim K, D_A \sim D_B \sim D$, allows to choose the characteristic time scale as $1/k$ and characteristic length scale as the characteristic distance on which protein diffuses over the characteristic time scale, $\sqrt{D/k}$. Then, looking at the equations, one can immediately conclude that the gradient, if it exists at all, can be maintained only over distances of the order of $\sqrt{D/k}$. Crude measurements of the diffusion coefficients and reaction rates can be made, and one can estimate if the chemical gradient of this sort can feasibly be relevant to the studied phenomenon.

5. Monte Carlo Simulations

Introduction into simple Monte Carlo simulations can be found in (Mogilner et al., 2002). Often, mathematical analysis is not sufficient and further computational tools are necessary. A lot can be achieved by simulating assumed rules underlining an investigated phenomenon. Very simple algorithm that is very straightforward to implement is shown in Fig. 3 of the main text. It is written as a Matlab code that works as follows:
1. Choose the initial position of the motor, set the step size and probability to take a step per second.
2. At each computational step, generate a random number and use it (and a simple procedure from the probability theory) to determine whether the motor takes a step or not.
3. Update the motor's trajectory.
4. Repeat the procedure until required.

The result can be plotted and compared visually with observed motor trajectories, or it is easy to compute the average velocity and effective diffusion coefficient of the motor and compare those with observables. Similar codes for complex models and respective simulations can be very long and involved.

6. Software for Modeling

The increase in use of quantitative modeling in cell biology is accompanied by a similar surge of software development. We list here but a few examples of the available tools from the entire available spectrum, from very multi-purpose ones that require both programming and math background to specific-purpose ones that were designed to be used by biologists without any math or programming experience. All software tools mentioned here are free of charge unless specified differently.

<u>General purpose numerical and symbolic packages</u>
Many numerical packages are being used in physical science and engineering communities for many years. These are general-purpose software tools that are very

versatile and can be used for almost any type of numerical and mathematical modeling needs. They usually require some knowledge of programming and mathematics.
Matlab - www.mathworks.com (Commercial)
SciLab - www.scilab.org
Octave - www.gnu.org/software/octave
Mathematica - www.wolfram.com (Commercial)

ODE solvers
These packages solve ODEs numerically and can be used to simulate models that are implemented using ODEs. No programming skills are required, but the user has to write up the equations themselves.
Xppaut – www.math.pitt.edu/~bard/xpp/xpp.html
Berkeley madonna – www.berkeleymadonna.com

PDE solvers
Virtual Cell – www.nrcam.uchc.edu – is invaluable resource.

Biochemical / genetic networks format
Many software tool specifically written to model and simulate biochemical and genetic networks were developed in the recent years. To allow software tools to communicate with one another a few standards emerged, the leading one being SBML (Systems Biology Markup Language sbml.org). Over 100 different free software tools support SBML and provide different functionalities: validation, simulations, libraries for other programming languages, network layout editors and format converter just to name a few. The combinations of those tools allow the user to design and simulate a biochemical or genetic network without any need for programming or mathematical expertise. A few examples of such tools are:

Systems Biology Workbench - sbw.kgi.edu
CellML – www.cellml.org
CellDesigner – celldesigner.org
BioSpice – biospice.org
Virtual Cell – www.nrcam.uchc.edu

Model Repositories
Many of the software tools mentioned above provide a repository of models implemented using respective package. An additional initiative that is unrelated to a specific software is the BioModels database (biomodels.net/) that currently holds many published models in a variety of different formats.

Supplemental Examples

1. Model of Flagellar Length Control (Marshall and Rosenbaum, 2001)

Flagella and cilia are organelles consisting of nine doublet microtubules projecting out from the cell surface (Mitchell, 2004). Flagella provide an interesting test-bed for studies of organelle size control, because they are easy to visualize, non-essential, and most importantly, their size can be described with a single number, the length. Within a given cell type, flagellar length tends to be quite fixed. Moreover, in a cell with more than one flagellum, all the flagella are of the same length. Flagellar length appears to be actively controlled, because a severed flagellum will rapidly re-grow to the correct length (Marshall, 2004). When one flagellum is severed and begins to re-grow, the other flagella on the cell respond by shortening, as though they "know" that they are longer than the recently severed one. The molecular mechanisms of length control are not known.

Mathematical models for length control have been devised for two purposes. One is to figure out how much the cell has to "know", that is, must there be a signal transduction mechanism that senses length, or can simpler mechanisms suffice. The other purpose is to provide clues about the types of molecular activities required for length control, as a step towards identifying the molecular players.

One reasonable model for length control would simply be that the length is determined by a fixed quantity of some structural protein. The cell could thus increase or decrease flagellar length by upregulating or downregulating the expression of the gene encoding this protein. Jarvik and co-workers developed an elegant approach for testing this mechanism, by employing a very simple mathematical model (Kuchka and Jarvik, 1982). Their approach is based on mutants of the green alga Chlamydomonas in which cells have a variable number of flagella, from zero to six per cell. They reasoned that if length was determined by the complete incorporation of some limiting precursor, then whatever number of flagella were present in a cell, the total length of all flagella should always add up to the same number, so if one were to measure flagellar length versus the number of flagella present, the resulting curve should show a geometric dependence of length on number. When Kuchka and Jarvik measured lengths in their variable-number mutant, they found that in fact length did not show such a geometric dependence on number but instead varied hardly at all over a range of 1-4 flagella. This result did away with the limiting-precursor model in a single stroke, without having to identify any of the molecules that might have been involved.

An alternative model for length control has recently been proposed (Marshall and Rosenbaum, 2001) based on the fact that flagellar microtubules undergo continuous turnover. Experimental studies revealed that tubulin is continuously removed from the distal tip of flagella, at a rate which is independent of length, and that this disassembly is normally balanced by continuous assembly which requires kinesin-mediated transport whereby large protein complexes called IFT particles move from the cell body out to the tip of the flagellum, carrying cargo proteins such as tubulin (Fig. S1A). The number of IFT particles within a flagellum was found to be independent of the length of the

flagellum. If a fixed number of transport complexes are forced to travel a greater and greater distance at a constant rate (Marshall et al., 2001), as the flagellum elongates, the round-trip transit time per particle will increase proportional to the length, *L*, and hence the overall rate of protein transport by the IFT particles will decrease as 1/*L*. Assuming assembly is transport limited, as suggested by the experiment (Marshall et al., 2001), the rate of assembly should be proportional to 1/*L*. In order to maintain a fixed length, the length-dependent assembly rate must equal the length-independent disassembly rate (Fig. S1B). Because assembly rate is a decreasing function of length, there is a single stable length at which the assembly rate equals the disassembly rate (Fig. S1B).

The rate of change of the flagellar length (*L*) can be characterized by the ordinary differential equation: $\frac{dL}{dt} = \left(\frac{I \cdot V}{2L} \right) K \left(P - NL \right) - D$, where *D* is the constant depolymerization rate, *I* is the number of IFT particles per flagellum, *V* is the speed of IFT particle, *P* is the total pool of flagellar material measured in units of length, *N* is the number of flagella, and *K* is the proportionality coefficient for the fraction of the pool of flagellar material associating with IFT particle. Note that $\left(P - NL \right)$ is the available pool of flagellar material, while *2L/V* is the time needed for one IFT particle to go back-and-forth along the flagellum, so *IV/2L* is the effective rate of IFT particles 'visiting' the flagellar tip, and $K \left(P - NL \right)$ is the length increment per one such visit. 'Lumping' the model parameters: $A = \frac{IVKP}{2}, B = \frac{IVK}{2}$, simplifies the equation: $\frac{dL}{dt} = \frac{A}{L} - \left(BN + D \right)$. It can be easily solved, either analytically, or numerically, predicting as a result the growth kinetics (Fig. S1C). Equating the right hand side of the equation to zero, allows finding the steady state length as the decreasing function of the number of the flagella: $L_{ss} = \frac{A}{BN + D}$ (Fig. S1D). Remarkably, all model parameters except two – *K* and *P* – can be measured, and then fitting the data to the theoretical curves in Fig. S1C,D gives the estimates for *K* and *P*. Then, the model can be truly tested (as there are no more free parameters) by measuring length kinetics of *M* flagella with various initial lengths and comparing the data to predictions of the generalized model, in which the dynamics of the i^{th} flagellum is described by the equation: $\frac{dL_i}{dt} = \left(\frac{I \cdot V}{2L_i} \right) K \left(P - \sum_{i=1}^{N} L_i \right) - D$.

This simple model can account for all reported phenomenological studies of length control, including the ability of flagella to equalize their lengths when one is severed (Marshall and Rosenbaum, 2001; Fig. S1C-D). The model also provides an excellent fit to measurements of flagellar length versus number (Marshall et al., 2005).

One important role of mathematical models in cell biology is to point out potentially key regulatory steps, in order to focus on the essential molecules. In the case of flagellar length control, the described balance-point model indicates that two key molecular players would be (a) whatever molecules set the number of IFT particles active within a flagellum, and (b) whatever molecules catalyze the continuous turnover of flagellar

microtubules at the distal tip. Now that the existence of these key players has been suggested by theoretical analysis, genetic and biochemical approaches can be devised to determine their molecular identities.

2. Microtubule-Dependent Centering (Kimura and Onami, 2005; Malikov et al., 2005)

One of the important examples of spatial self-organization of the cell is an ability of cytoskeletal structures to find geometric center of the cell. Often, respective processes are based on dynamics of microtubule asters and associated molecular motors. These processes were extensively studied, both experimentally and theoretically, in vivo and in vitro, revealing an amazing diversity of the underlying molecular mechanisms (Reinsch and Gönczy, 1998; Burakov et al., 2003; Tran et al., 2001; Holy et al., 1997; Vallee and Stehman, 2005). Here we review two recent studies that combined live imaging and computer modeling to elucidate microtubule-dependent centering machinery (Kimura and Onami, 2005; Malikov et al., 2005).

In C. elegans, male and female pronuclei migrate from the periphery toward the center of the egg following fertilization. This migration depends on microtubules growing from two centrosomes associated with the male pronucleus. Two mechanisms – "pushing" and "pulling" – were previously proposed for this migration (Reinsch and Gönczy, 1998) (Fig. S2A). In the pushing mechanism, plus ends of the microtubules that reach the cell edge continue to polymerize against the boundary. This growth generates a polymerization ratchet force (Dogterom and Yurke, 1997; Mogilner and Oster, 1999) that pushes the unyielding cell boundary and, as a reaction, buckles these microtubules (Fig. S2A). Mechanically, microtubules are elastic rods, for which the buckling force is inversely proportional to their length squared (Howard, 2001), so when the pronucleus is closer to the right edge of the cell (Fig. S2A), the short buckling microtubules at the right push the pronucleus to the left. The corresponding force imbalance is great, and so is the resulting migration speed proportional to the force in the low Reynolds numbers environment of the cell. Closer to the cell center, microtubules both at the left and at the right reach the boundaries, but at the right they are still shorter than at the left, so the pushing force from the right is greater, and the migration continues, albeit at decreasing rate. Eventually, when the pronucleus is at the center, the forces balance and the migration stops.

On the contrary, the pulling mechanism assumes that minus-end-directed motors anchored throughout the cytoplasm (i.e., dynein molecules) pull the microtubules, and as a result pull the pronucleus. An important model assumption is that there is a constant number of motors per unit length of microtubules, and that the motor forces are additive, so the resulting pulling force is proportional to the microtubule length (Fig. S2A). According to this mechanism, when the pronucleus is close to the right edge of the cell, it is pulled to the left stronger than to the right, because microtubules growing to the right are restricted in length by the edge. However, the resulting force is small, because the microtubule aster initially consists only of short microtubules – they do not have time to grow significantly yet. Later, when the pronucleus is closer to the center, the misbalance

increases because the microtubules grow faster than the pronucleus migrates, and the migration accelerates. Closer to the center of the cell, the lengths of the left- and right-oriented microtubules equilibrates, and the migration slows down.

Kimura and Onami (2005) simulated both mechanisms using Monte Carlo method to keep track of every microtubule in the aster undergoing dynamic instability and respective forces and discovered that the resulting positions of the pronucleus as functions of time are qualitatively different for these two mechanisms in complete agreement with the qualitative arguments: a pushing mechanism predicts a hyperbolic function (rate of centering decreases with distance to the center), while a pulling mechanism predicts a sigmoidal function (the rate first increases, then decreases with distance) (Fig. S2C). Microscopy revealed that the observed function is indeed sigmoidal (Kimura and Onami, 2005), so this combination of modeling and experiment argues in favor of the pulling mechanism of centering.

The centering can be completely forceless in other systems (Malikov et al., 2005). For example, in microsurgically produced cytoplasmic fragments of fish melanophores (pigment cells), microtubules are initially oriented with their minus ends toward the cut edge (right in Fig. S2B) and plus ends toward the opposite (left) edge, reflecting their orientation in the "mother" cell. The pigment granules in these fragments are scattered throughout the cytoplasm and are coated with multiple molecular motors including minus-end-directed dyneins (Vorobjev et al., 2001). Upon stimulation of these motors, the granules rapidly aggregate to the right edge of the fragment, toward the minus ends of the microtubules "inherited" from the mother cell (Fig. S2B). Surprisingly, the aggregate of granules then slowly migrates to the center of the fragment (Fig. S2B). Experiments of Malikov et al. (2005) revealed that the contact of the microtubules with fragment's boundary is not necessary, so the pushing mechanism is out, as is the pulling mechanism, because the microtubules are immobile and cannot be pulled.

Microscopy suggested, and computer simulations confirmed that the peculiar geometric centering mechanism in this system is based on self-nucleated microtubules (as opposed to the majority of microtubules nucleated on granules) which act as a volume sensing tool (Fig. S2B). More such microtubules would nucleate on the side of the aggregate away from the nearest edge as the rate of spontaneous nucleation per unit of cytoplasmic area is assumed to be constant across the cytoplasm. This asymmetry leads to a bias in the transport of pigment granules in the aggregate and directs their movement away from the cut edge (Fig. S2B). When the aggregate reaches the center of the fragment, the cytoplasmic area and thus the probability of microtubule nucleation on all sides of the aggregate becomes equal, keeping the aggregate in the center, equidistant from the fragment's margins. Computer simulations allowed investigation of the centering dependence on parameters of microtubule dynamics which are impossible to vary experimentally, and predicted the centering kinetics (Fig. S2D), which turned out to be in an excellent agreement with the data. In the future, one of the greatest challenges in cell biology, which will be impossible to meet without modeling, is to understand how multiple redundant mechanisms, such as the described centering phenomena, cooperate and/or compete in live cells.

Supplemental References

Adler F. 2005 Modeling the Dynamics of Life: Calculus and Probability for Life Scientists. New York, Thomson.

Burakov A, Nadezhdina E, Slepchenko B, Rodionov V. Centrosome positioning in interphase cells. J Cell Biol. 2003 162:963-9.

Dogterom M, Yurke B. Measurement of the force-velocity relation for growing microtubules. Science. 1997 278:856-60.

Edelstein-Keshet L. 1988 Mathematical Models in Biology. New York, Random House.

Holy TE, Dogterom M, Yurke B, Leibler S. Assembly and positioning of microtubule asters in microfabricated chambers. Proc Natl Acad Sci U S A. 1997 94:6228-31.

Howard, J. 2001. Mechanics of motor proteins and the cytoskeleton. Sinauer, Sunderland, MA.

Kholodenko BN. Cell-signalling dynamics in time and space. Nat Rev Mol Cell Biol. 2006 7:165-76.

Kimura A, Onami S. Computer simulations and image processing reveal length-dependent pulling force as the primary mechanism for C. elegans male pronuclear migration. Dev Cell. 2005 8:765-75.

Kuchka, M.R., and Jarvik, J.W. 1982. Analysis of flagellar size control using a mutant of Chlamydomonas reinhardtii with a variable number of flagella. *J. Cell Biol.* 92,170-175.

Lin CC, Segel L. 1974 Mathematics Applied to Deterministic Problems in the Natural Sciences. New York, Macmillan.

Logan JD. 1977 Applied Mathematics. New York, John Wiley & Sons.

Malikov V., Cytrynbaum E.N., Kashina A., Mogilner A., Rodionov V. Centering of a radial microtubule array by translocation along microtubules spontaneously nucleated in the cytoplasm. Nat. Cell Biol. 2005.7:1213-8.

Marshall, W.F., and Rosenbaum, J.L. 2001. Intraflagellar transport balances continuous turnover of outer doublet microtubules: implications for flagellar length control. *J. Cell Biol.* 155,405-414.

Marshall, W.F., Qin, H., Brenni M.R., and Rosenbaum., J.L. 2005. Flagellar length control system: testing a simple model based on intraflagellar transport and turnover. *Mol. Biol. Cell* 16,270-278.

Marshall WF. Cellular length control systems. Annu Rev Cell Dev Biol. 2004 20:677-93.

Mitchell DR. Speculations on the evolution of 9+2 organelles and the role of central pair microtubules. Biol Cell. 2004 96:691-6.

Mogilner A, Oster, G. The polymerization ratchet model explains the force-velocity relation for growing microtubules. Eur. Biophys. J. 1999 28:235-242.

Mogilner, A., Elston, T., Wang, H.-Y. and Oster, G. 2002. Molecular motors: theory and examples, in *Computational Cell Biology* , C. P. Fall, E. Marland, J. Tyson and J. Wagner, Eds., pp. 321-384, Springer, N.Y.

Reinsch S, Gonczy P. Mechanisms of nuclear positioning. J Cell Sci. 1998 111:2283-95.

Tran PT, Marsh L, Doye V, Inoue S, Chang F. A mechanism for nuclear positioning in fission yeast based on microtubule pushing. J Cell Biol. 2001 153:397-411.

Vallee RB, Stehman SA. How dynein helps the cell find its center: a servomechanical model. Trends Cell Biol. 2005 15:288-94.

Vorobjev I, Malikov V, Rodionov V. Self-organization of a radial microtubule array by dynein-dependent nucleation of microtubules. Proc Natl Acad Sci U S A. 2001 98:10160-5.

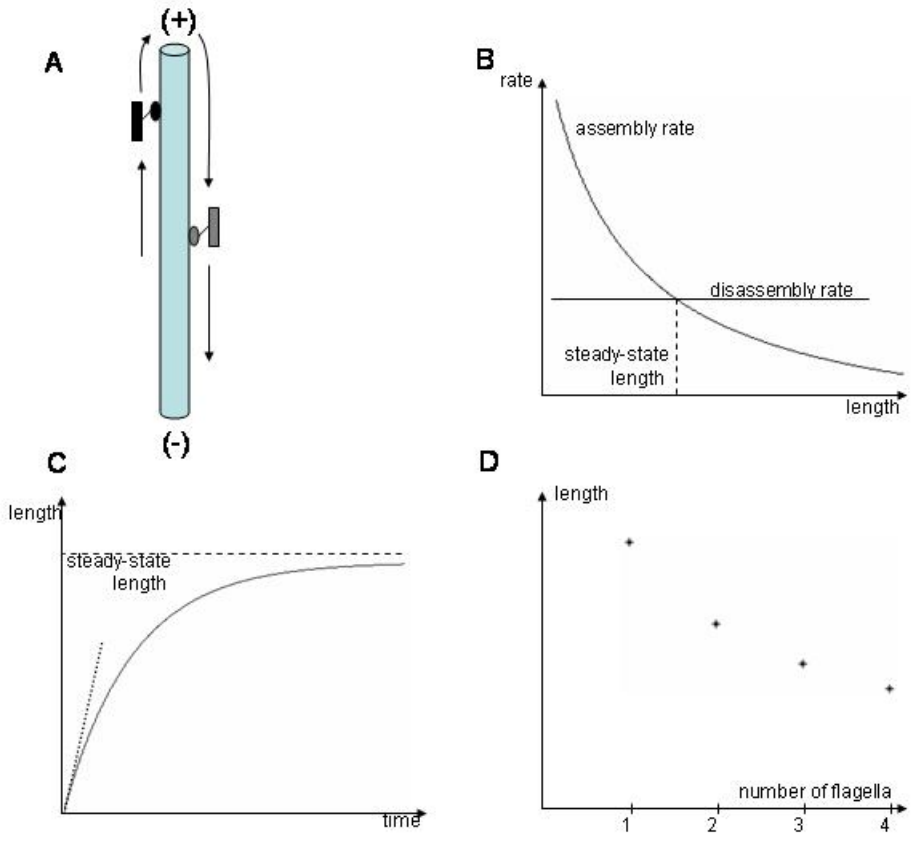

Figure S1. Flagellar Length Control

A. Plus-end-directed motor/IFT particle complexes deliver building material to the flagellar tip, while minus-end-directed motor/IFT particle complexes disassemble the tip at a constant rate. B. Balance-point model: the stable steady-state length of the flagellum is determined by the balance between the constant disassembly rate and the assembly rate which decreases in inverse proportion to the length. C. Kinetics of flagellar growth in the regeneration experiment (solid curve) predicted by the solution of the model equation. The dotted line – slope of the tangent to the solid curve at $t = 0$ – is the maximal growth rate. D. Dependence of the flagellar length on the number of flagella per cell predicted by the solution of the model equation.

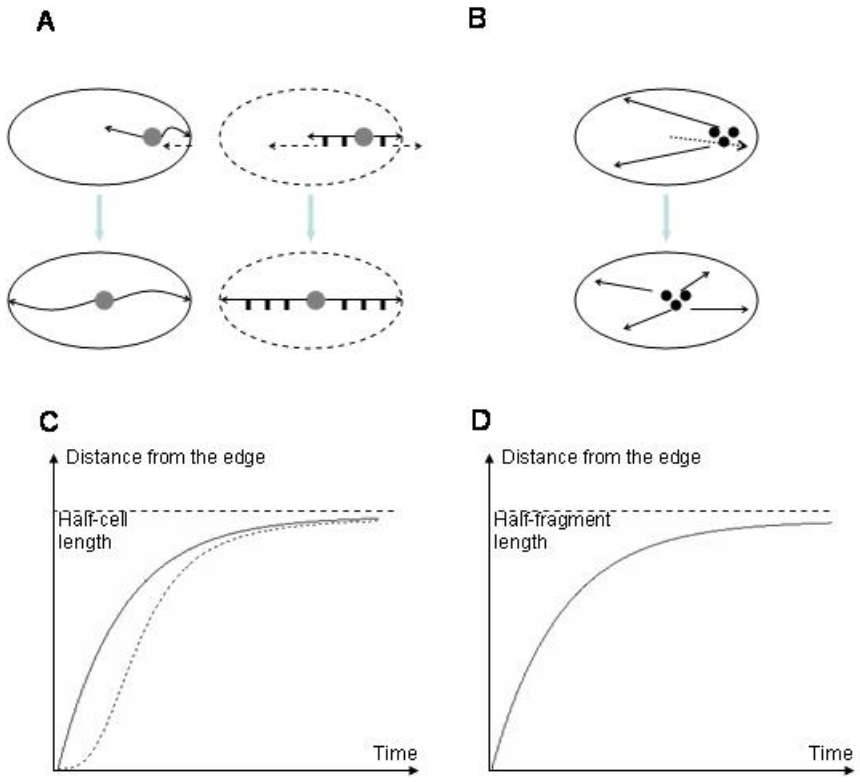

Figure S2. Microtubule-Dependent Centering

A. Force-dependent centering mechanisms. Left: "pushing" mechanism. Short microtubules (plus ends are indicated by the arrows) grow at the cell edge, buckle, and generate pushing force (dashed arrow) until the buckling forces are balanced when the pronucleus is at the cell center. Right: "pulling" mechanism. Minus-end directed motors (black rectangles), the number of which is proportional to microtubule length, generate forces (dashed arrows) that are unbalanced because microtubule growth to the right is restricted by the cell edge and, as a result, that are pulling the pronucleus toward the center, where the forces balance. B. Forceless centering mechanism: initially, minus-end directed motors transport pigment granules to the edge of the cell fragment. Then, self-nucleated microtubules (dotted) serve as tracks moving the granules toward the center. C-D. Computer simulations predict the centering kinetics. Importantly, the pulling mechanism leads to sigmoidal kinetics (dotted), unlike the hyperbolic kinetics of the pushing and forceless mechanisms.