“Essentially, all models are wrong, but some are useful”—a cross-disciplinary agenda for building useful models in cell biology and biophysics

Julien Berro^{1,2,3}

Received: 24 August 2018 / Accepted: 30 October 2018 / Published online: 12 November 2018

© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany, part of Springer Nature. 2018

Abstract

Intuition alone often fails to decipher the mechanisms underlying the experimental data in Cell Biology and Biophysics, and mathematical modeling has become a critical tool in these fields. However, mathematical modeling is not as widespread as it could be, because experimentalists and modelers often have difficulties communicating with each other, and are not always on the same page about what a model can or should achieve. Here, we present a framework to develop models that increase the understanding of the mechanisms underlying one’s favorite biological system. Development of the most insightful models starts with identifying a good biological question in light of what is known and unknown in the field, and determining the proper level of details that are sufficient to address this question. The model should aim not only to explain already available data, but also to make predictions that can be experimentally tested. We hope that both experimentalists and modelers who are driven by mechanistic questions will find these guidelines useful to develop models with maximum impact in their field.

Keywords Model · Theory · Cross-disciplinary research

Introduction

A popular joke about modeling goes as follows: A group of farmers desperately trying to increase their cows’ milk production calls a theorist to help them find a solution. After a few months of hard work, the theorist calls back: “I found the optimal solution. Consider a spherical cow in a vacuum ...”

The trove of quantitative data produced by modern biology has highlighted that the complex behaviors of biological systems, even the simplest ones, are difficult to comprehend with our intuition alone (Mogilner et al. 2006; Pollard 2010; Pollard and De La Cruz 2013; Howard 2014; Marshall 2017). An increasing number of experimentalists appreciate the need for mathematical modeling to explain their data and

uncover the underlying molecular mechanisms for their favorite biological systems (Pollard 2013; O’Shaughnessy and Pollard 2016). However, mathematical modeling is not as widespread as it could be. We believe one of the main reasons for the limited use of modeling by experimentalists in the cell biology and biophysics communities is a common misunderstanding of what mathematical modeling can achieve. As the spherical cow joke demonstrates, another common reason is that mathematicians and physicists who model biological processes sometimes make simplistic assumptions, or do not explain or justify the simplifications clearly enough when they are legitimate. Last, good communication between fields is important to develop models, and produce outputs that can be translated by experimentalists into testable predictions.

Joke aside, what would be the best way to model a cow if a sphere is not a good model? Should we include its limbs, head, and tail? Its digestive system? The different cell types involved in digestion? The dynamics of actin filaments in each cell? The atomic interactions between the actin subunits? One could enumerate infinite details but it is clear that not all details are meaningful or relevant for the problem at hand. Indeed, the appropriate level of details for the model depends on the question asked. A sphere is probably the best model if one wants to know how far to catapult a cow, like in the classic

✉ Julien Berro
julien.berro@yale.edu

¹ Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA

² Department of Cell Biology, Yale University School of Medicine, New Haven, CT, USA

³ Nanobiology Institute, Yale University, West Haven, CT, USA

“Monty Python and the Holy Grail” movie (Gilliam and Jones 1975). Even a point with a mass would be a sufficient model to calculate a rough estimate. Likewise, if one wants to know how long it takes to freeze a cow in liquid nitrogen, a sphere would be an excellent first model. To answer the farmers’ milk productivity question, the cow’s shape and dimensions are unimportant details, but a deep qualitative and quantitative understanding of the cow’s physiology are critical. After some research, one may even realize that too little is known about the physiology of milk production, and trying to answer the farmers’ question with mathematical modeling might not be a good idea after all.

These examples are perfect illustrations of the statistician George EP Box’s quote: “essentially, all models are wrong, but some are useful” (Box and Draper 1987). The goal of a mathematical model in biology is not to be a complete, explicit representation of the biological system that is as detailed as the underlying biochemistry (Mogilner et al. 2006; Möbius and Laan 2015) but to be “useful” in increasing our understanding of the mechanisms of this biological system (Mogilner et al. 2012; Pollard 2013). In this paper, we propose guidelines for developing “useful” models in cell biology and biophysics. We hope that both experimentalists and modelers who are driven by mechanistic questions will find useful guidelines to develop models with maximum impact in their field. This paper also aims to catalyze collaborations between experimentalists and modelers, by helping them to find a common ground to build models, to figure out what is possible to achieve and what is unrealistic, to learn what to expect during the development of the model and where the hurdles are, and in general how to get the most out of their modeling efforts.

What do we mean by “useful” models?

What is a model?

A model is a simplified representation of the biological system studied, following simple rules that can be formalized into mathematical equations or into a computational algorithm. One should see the model as a caricature of the system studied rather than its exhaustive, detailed representation. A model is in its essence not an exact and complete representation of the reality, nor should it be one, if one wants to keep it more tractable than the biological system studied (Mogilner et al. 2006; Howard 2014; Möbius and Laan 2015). In that sense, all models are wrong at least in some of their details, but their overall direction is very likely to be correct, just like the conceptual models and cartoons one draws to represent their favorite biological system (Rothman 2018).

There are two main classes of mathematical models: computational models and theory. A computational model (or simulation) is a set of explicit rules, equations, or

algorithms that aim to reproduce a simplified representation of the system. To some extent, computational models can be considered as experiments performed *in silico* (i.e., with a computer), as they can be used in the same way one performs an *in vitro* or *in vivo* experiment. They often are relatively easier to implement by non-specialists than theory, especially using generic tools and software packages that are readily available (see below).

Theory (or theoretical models) is a set of mathematical equations that describe the biological system. Theory usually aims to explicitly determine how key properties of the biological system change with parameter values or to identify critical parameters and relationships between parameters which define different regimes, by deriving formulas from the model equations, without running simulations (Howard 2014; Phillips 2015, 2017). However, such closed formulas are usually very difficult (or even impossible) to obtain except in specific cases of models that have a low number of variables and parameters, or at the cost of certain simplifications. There are no systematic, straightforward methods to develop theoretical models, and they usually require very good physical and mathematical skills and intuition, and a lot of effort to derive.

What are models used for?

Models are usually used to determine whether the mechanisms one hypothesizes for their favorite biological system can reproduce the observed data quantitatively, or at least qualitatively. A mathematical model is actually the best way to determine whether one’s favorite cartoon model is indeed valid. In addition, mathematical models may allow one to estimate the values of unknown parameters by fitting the model to the data. Models are also used to explore alternative mechanisms, to determine in which conditions the assumptions are valid, or to identify what components or interactions are sufficient to explain the data.

Even if experimental data are not available, one can use models to gain a better sense of the overall output or emergent properties of multiple simple interactions, for which intuition is not sufficient (Pollard 2010; Mogilner et al. 2012; Pollard and De La Cruz 2013; Gunawardena 2014; Goldstein 2018). One can also better understand how the system behaves when parameters are varied, which may help design experiments to challenge the hypotheses of the model. Last but not least, theoretical models allow one to derive formulas that connect variables and parameters of the model to quantities that are experimentally measurable, and allow one to determine the different regimes of the system. They also allow one to generalize properties and determine which parameters or combination of parameters are the most meaningful for the system.

What makes a model useful?

A useful model is a model that advances our mechanistic understanding of the biological process studied. In that sense, even a model that does not quantitatively fit experimental data can be useful, since it demonstrates that the proposed mechanisms are incomplete or incorrect, and forces one to rethink or complete the proposed mechanisms. In addition, a model that does not have a direct application or for which experimental data are not available can also be useful as long as it enlightens us on possible and impossible mechanisms or outcomes (Flexner 1939). A useful model also allows one to make predictions that can be experimentally tested. These predictions should aim to further validate one's hypotheses, invalidate alternative hypotheses, or experimentally demonstrate new features in the biological process studied.

The fact that one can find a model and some set of parameters that accurately fits the data does not prove that the hypotheses underlying the model are true, but only proves that they are compatible with the observed data. In fact, alternative mechanisms could also lead to the same output, and models should aim to also demonstrate that common alternative hypotheses cannot reproduce the experimental data, or help design experiments that can distinguish between different possible mechanisms.

Tentative agenda to develop useful models

In the following sections, we will present nine key steps to develop useful models, which will be illustrated with examples from two models published by the Pollard lab. The first model, referred to as the cytokinetic ring model (Fig. 1), aimed to uncover the molecular mechanisms of assembly of the contractile ring at the beginning of cytokinesis in fission yeast (Vavylonis et al. 2008). The second model, referred to as the endocytic actin model (Fig. 2), aimed at determining the molecular mechanisms underlying fast actin dynamics during clathrin-mediated endocytosis in yeast (Berro et al. 2010b). Note that both models are computational models but our proposed agenda could also be applied when developing theoretical models.

1. Find a good biological question

Before developing any model, it is critical to clearly identify what questions the model aims to answer and which hypotheses or mechanisms it aims to test. As we saw with the cow models, the question addressed by the model strongly determines the modeling approach and what level of detail is appropriate.

In general, one should avoid questions that have already been extensively addressed with mathematical modeling,

Fig. 1 The cytokinetic ring model (Vavylonis et al. 2008). Data: Previous experiments showed that the cytokinetic contractile ring assembles from the condensation of nodes into a ring (micrographs: myosin regulatory light chain Rlc1p-3GFP marks nodes at progressive stages of ring formation). Question: The model aimed to determine which mechanisms allow the condensation of a broad band of nodes into a homogenous contractile ring. Model: The initial hypothesis was that actin filaments (green) nucleated from the nodes (red) extend in random directions (search), nodes near a filament attach to it (capture), and exert force (pull). The alternative mechanism included the same assumptions and added a random breakage of the links between the nodes and the filaments (release). Results: The initial hypothesis led to the formation

of clumps of nodes, whereas the alternative model aligned the nodes better and the ring was more regular. Model validation: The timing of node alignment, which was not explicitly included in the assumptions, was the same as in the experiments. The model also reproduced the condensation of rings from known mutants with very broad bands or double rings. Predictions: A broad parameter screen and sensitivity analysis made many predictions, including that reduction of the release rate should lead to less homogenous rings. Experimental tests: Fission yeast cells expressing ADF/cofilin mutants with lower severing activity make inhomogeneous rings with clumps of nodes (Chen and Pollard 2011). Images adapted from (Vavylonis et al. 2008). Reprinted with permission from AAAS

Fig. 2 The endocytic actin model (Berro et al. 2010b). Data: Previous experiments showed that actin, the Arp2/3 complex, its activator WASP, and many actin-binding proteins are recruited to sites of clathrin-mediated endocytosis in fission yeast. The temporal evolution of the copy number of 12 proteins was recently measured (micrograph: temporal evolution of fim1p-GFP at one endocytic site). Question: The model aimed to test whether dendritic nucleation by the Arp2/3 complex was able to reproduce the fast assembly and disassembly of actin during endocytosis. Model: The initial model hypothesized that dendritic nucleation and depolymerization from the pointed ends created via severing and debranching were sufficient to explain the fast actin dynamics. The model was largely constrained by rate constants that had been previously measured in vitro. Failure of this model to reproduce fast disassembly led us to hypothesize that the release of chunks of the actin

meshwork increases disassembly. Results: Outputs of the initial model (solid lines) did not match the experimental data (symbols), whereas the alternative model was able to reproduce the data (teal: actin; black: Arp2/3 complex; red: WASP; green: capping protein). Adapted from (Berro et al. 2010b). Model validation: Another model where depolymerization rate was 2 orders of magnitude larger than ever measured could also reproduce the experimental data but this model was not robust to small changes in parameter values, and was therefore rejected. Predictions: The model predicted that reduced severing activity has large effects on protein disassembly but small effects in protein assembly. Experimental tests: This prediction was validated using a strain expressing an ADF/cofilin mutant with reduced severing activity (Chen and Pollard 2013)

except if one has a new original hypothesis to test or if new experimental data are available. Note that not all biological questions can or should be addressed using mathematical modeling. It is sometimes better to perform new well-chosen experiments rather than to develop a model, especially when the amount of information on the biological system is very sparse and not sufficient to distinguish between hypotheses.

The cytokinetic ring and endocytic actin models both aimed at answering specific mechanistic hypotheses. Before developing the cytokinetic ring model (Vavylonis et al. 2008), it had been observed that in fission yeast, individual diffraction-limited spots called nodes are present in a broad band around the cell equator before the onset of cytokinesis, and these nodes slowly realign into a thin, homogeneous contractile ring, followed by ring constriction and cell separation (Wu et al. 2006). However, the mechanisms driving the node condensation from a broad band into a thin, homogenous ring remained elusive and an important question in the field. The presence of actin filament-nucleating formins and myosin II motors within every node led Vavylonis et al. to hypothesize

that actin filaments nucleated from one node may come in close proximity with another node, be captured by its myosins, consequently pulling both nodes toward each other and eventually leading to the formation of a ring. In addition, the authors had recently made key quantitative measurements, including a characterization of the node movements during the formation of the contractile ring. The model aimed at testing whether this “search-capture-pull” mechanism was able to explain the observed data.

When the model for actin dynamics during clathrin-mediated endocytosis was developed, the field had already identified that ~20 proteins involved in actin assembly and disassembly are recruited to endocytic actin patches in yeast (Kaksonen et al. 2003, 2005, 2006; Sirotkin et al. 2005). Among them were the Arp2/3 complex, WASP, and the canonical actin capping protein. Quantitative microscopy had also recently determined the temporal evolution of the copy number of these proteins during their assembly to and disassembly from the endocytic structures (Sirotkin et al. 2010; Berro and Lacy 2018). The common hypothesis in the field

was that actin assembly was driven by autocatalytic dendritic nucleation of actin filaments, and the creation of new pointed ends via the severing of actin filaments by ADF/cofilin enhanced the disassembly by depolymerization of the actin in the endocytic patch. The model aimed at testing whether these hypotheses could quantitatively reproduce the very fast assembly and disassembly of actin during endocytosis observed in the experiments.

2. Determine the most appropriate modeling approach

Overarching principles The most important overarching principle when building a model is to keep it as simple and as coarse as possible, retaining the details that matter most for the biological question, and, only if necessary, refining the model one assumption at a time. It is also important to use the empirical principle of Occam's razor, or law of parsimony, which states that between competing models that match the same data, one should keep the one with the fewest assumptions and unknown parameters. Indeed, models with the fewest assumptions allow to capture the mechanisms that have the largest effects. Counterintuitively, the simplest models are often the most predictive, and the most useful. In fact, the mathematician John von Neumann used to say, “with four parameters I can fit an elephant, and with five I can make it wiggle its trunk” (Dyson 2004), which is indeed possible (Mayer et al. 2010).

Another important principle is to ensure that the property that one tries to explain with the model is not already included in its assumptions. For example, a model for the contractile ring that would include a force field explicitly driving the nodes toward the center of the cell would not be very useful if one wants to understand the mechanisms underlying the emergence of a thin ring from a broad band of nodes.

There is usually more than one way to develop a model that addresses a specific biological question. It often requires good intuition and several iterations to identify the appropriate level of detail and the best approach. A lot can be learned just going through the process of building the model. Indeed, developing a model forces one to specify each aspect of the model, and often makes one realize that some details of the system are much less well understood than commonly assumed, or have been consciously or unconsciously swept under the rug by the field because nobody really dared opening the can of worms (Berro et al. 2010a; Wick and Kane 2011; Gunawardena 2014). For example, while developing the endocytic actin model, we realized that the fate of WASP after it activates an Arp2/3 complex had been largely ignored. Does WASP remain attached to the membrane? Does it detach from the membrane? Does it remain active and ready to activate another Arp2/3 complex?

What is known about the mechanisms and what is the quality of this information? In general, other researchers in the field have already proposed mechanisms for the system one tries to model. One should not hesitate to re-evaluate the published experimental and modeling data in light of these hypotheses, since authors often favor their preferred hypothesis to explain their data, and new information might have been published since then. This step requires critical (re-)reading of the literature, and very good understanding of the different modeling and experimental techniques used in the papers. This step can be quite difficult for someone who has not used these techniques before, but one should not hesitate to reach out to collaborators or consultants for advice.

During the development of the endocytic actin model, the dendritic nucleation hypothesis, i.e., the auto-catalytic assembly of actin driven by the creation of new branches on existing filaments with the Arp2/3 complex, was well supported by previous genetics and microscopy data (Chang et al. 2003; Kaksonen et al. 2003, 2005; Sirotkin et al. 2005; Gachet and Hyams 2005; Kim et al. 2006). However, recruitment and activation mechanisms for WASP were unclear. Some pieces of information about the spatial organization of endocytic proteins had just been published (Idrissi et al. 2008), but their regulation by membrane geometry and force production mechanisms were virtually unknown. The rate constants of the Arp2/3 activation pathway had been systematically measured with purified proteins *in vitro* by Beltzner and Pollard (2008). However, in re-reading this paper, it became clear that some parameters might be different in the context of yeast endocytosis, because measurements were done with purified proteins from different species and some parameter estimates lacked precision.

The cytokinetic ring model was built from the well-documented existence of nodes that contain formins and type II myosins (Chang et al. 1997; Motegi et al. 2000; Wu et al. 2003, 2006; Lord et al. 2005), each protein having been well studied *in vitro* and *in vivo* (Naqvi et al. 2000; Kovar et al. 2003, 2005; Matsumura 2005). Exact reaction rates were not known, but new experimental characterization of node dynamics and ring formation had been recently obtained in the Pollard lab (Vavylonis et al. 2008). The observation that mutant fission yeast cells having a much broader node distribution than wild-type cells form two cytokinetic rings instead of one was a great opportunity to test the model.

What is the appropriate scale and granularity for the model?

The level of detail of the model closely depends on the question asked and the experimental data available. If the model is more detailed than the available data, there is a high risk to make many arbitrary choices and speculations, and the model will have too many unknown parameters (Phillips 2017). Conversely, if the model is too coarse, it may miss features that are important to answer the biological question asked. A

good starting point is to develop a model with details one level finer than the emerging property the model aims to capture, and refine if necessary. For example, nodes and actin filaments are the largest entities that constitute a cytokinetic ring, and were chosen as the minimum entities for the model.

Another choice to make is whether the model should include space and time components. The dimensions of space should be kept as small as possible to minimize the number of variables and parameters, and models with low dimensionality are more computationally and mathematically tractable. To keep the space dimensions as small as possible, one should determine which space information is critical, and take advantage of the symmetries of the system. When possible and relevant, one can remove the time component by assuming the system is at steady state.

When the cytokinetic ring model was implemented, it was known that the nodes and the actin filaments of the cytokinetic ring are essentially constrained to a thin region close to the plasma membrane. Modeling ring formation on a 2D cylindrical surface rather than a 3D volume was an excellent first approximation. This approximation greatly reduced the complexity of the model, the number of unknown parameters, and the simulation times.

At the time the model for endocytic actin dynamics was developed, the protein organization and movements within the endocytic structure were poorly understood, and not critical to the question the model aimed to address. Building a spatial model would have led us to many speculations and unknown parameters, which would have greatly reduced the model's predictive power. Therefore, we developed a model only taking into account time, implicitly assuming the proteins in the endocytic structure were homogeneously distributed (or “well mixed”).

Should the model be continuous or discrete? A discrete approach takes into account individual entities. A continuous approach considers that the system has enough entities and is homogeneous enough that the behavior of individual entities is not critical, and monitoring densities or concentrations is sufficient to capture the overall behavior of the system. For example, in the cytokinetic ring model, individual nodes were considered discretely because experimental data showed clearly separated entities that seemed to behave independently from each other. In the endocytic actin model, the number of molecules was high enough; therefore, the model ignored spatial organization and assumed proteins were “well mixed,” which allowed us to monitor overall concentrations as continuous variables rather than monitoring individual molecules.

Should the model be deterministic or stochastic? In a deterministic model, entities evolve according to determined rules, and for a given state of the system at an instant t there is a unique possible outcome at time $t + dt$. For example, at each time step dt , the amount of polymerized actin increases by

$k_+[BE][G - actin]dt$, where k_+ is the polymerization rate constant, $[BE]$ and $[G - actin]$ the concentrations of barbed ends and monomers respectively. In a stochastic model, the fate of the model entities at each time step depends on probability distributions, and several iterations of the same model in the same state produce different outcomes. For example, in the cytokinetic ring model, the direction of elongation of a new filament nucleated from a node was random, each direction having the same probability. In general, continuous models follow deterministic rules, and discrete models follow stochastic rules or a mix of deterministic and stochastic rules.

Should one use computational simulations or theory?

Theories are usually more difficult to develop than simulations. Very good mathematical and physical skills and intuition are often needed to develop valid theories and to identify regimes or conditions where the number of parameters and variables can be reduced. However, when a theory can be developed, it usually yields deep results, such as explicit relationships that link the macroscopic measurable properties of the system with the microscopic variables and parameters. For example, theory allowed Leonor Michaelis and Maud Menten to derive the famous Michaelis-Menten formula that links the rate of product formation during an enzymatic process to the different rate constants of the intermediate binding and catalytic reactions (Michaelis and Menten 1913).

Simulations are in general easier to implement by non-specialists, and broadly applicable user-friendly tools have been developed (see below). Simulations can reveal valuable insights on their own, such as constraints on parameter values and empirical relationships between parameters and variables. Note that non-standard problems can be complicated to implement and require good coding skills.

For systems that seem too complex to be easily tractable with theory at first sight, a successful strategy is to start with simulations to gain a better sense of the system's behavior and to identify conditions where simplifications can be made, making the systems more easily tractable with theory (Roland et al. 2008).

3. Implement the model

Simple theories can often be worked out with just a pen and paper, and one can use symbolic mathematics software such as Wolfram Mathematica or Maple to derive the equations. Specific numerical solutions can also be calculated and plotted with these tools or with other scientific numerical software such as MATLAB, Python, or R to name a few popular ones.

Simulations of systems of biomolecular interactions can often be performed with one of the many software packages that have been developed for a general use, and generally include an easy-to-use graphical user interfaces (reviewed in (Gostner et al. 2014; Thomas and Schwartz 2017)). Some

software are specialized to specific applications or modeling approaches, such as MATLAB's Simbiology plugin, COPASI (Hoops et al. 2006) and PySB (Lopez et al. 2013) for non-spatial biochemical kinetics, or Smoldyn (Andrews and Wren 2017) for Brownian dynamics of molecule interactions to name a few. Other software packages, such as Virtual Cell (Slepchenko et al. 2002; Blinov et al. 2017), provide a broad range of approaches and applications. Interoperability between different modeling simulation tools has been made easier thanks to the development of standards for model descriptions such as SBML (Hucka et al. 2003) and rule-based modeling languages such as BioNetGen (Harris et al. 2016) and Kappa (Boutillier et al. 2018).

For more complex models, one generally has to develop their own tools from scratch. This requires excellent math and coding skills, is usually quite time-consuming, and does not guarantee the implementation is accurate.

Whether one uses theory or simulations, best practice is to identify good “controls” to ensure that the model is implemented properly. For example, one can simulate a subset of the reactions where the outcome is known or easier to predict, or simplify theories in limit cases where one or several model variables vanish or become very large relatively to others.

The endocytic actin model, which implemented non-spatial kinetics, was implemented with MATLAB's Simbiology plugin, which allowed for easy scripting of optimization tasks, sensitivity analysis, and plotting. The cytokinetic ring model required custom code development in Java.

4. Identify model parameters that fit the data

Once the model is implemented, the next step is to make the model output fit the experimental data. In general, several parameters of the model are unknown and need to be determined. One can screen parameters until the model fits the data well enough, or one can perform an automatic fit of the parameters to the data (also called optimization), if the software used to implement the model has this functionality. Note that screening for parameters that fit the data should be done with caution because it does not guarantee that the parameter set that seems to best fit the data is actually the best fit, or the only set that does—other parameter sets might lead to equally good if not better fits. In addition, one should keep in mind that experimental data have limited precision and accuracy and one should avoid overfitting the model parameters with a precision better than the precision of the experimental measurements.

One may be disappointed if the first draft of the model cannot fit the data. However, this is a blessing in disguise since it demonstrates that the proposed mechanisms can be rejected (Popper 1959), or require refinement.

The “search-capture-pull” mechanism initially proposed for the cytokinetic ring assembly was able to align the nodes, but produced clumps instead of evenly spaced nodes, for all parameter sets tested. This discrepancy led the authors to revisit their model and realize that the regular breakage of links between nodes was required to obtain an even node distribution. This prediction led the authors to have a closer look at their experimental data and notice that this release was indeed happening.

The initial endocytic actin model was constrained with parameters that had been previously measured *in vitro* (Blanchoin and Pollard 1999; Fujiwara et al. 2007; Kuhn and Pollard 2007; Beltzner and Pollard 2008). However, this initial model was unable to reproduce the fast actin assembly and disassembly observed experimentally. In particular, we realized that fast disassembly could not be produced by pointed end depolymerization alone, except if unrealistic depolymerization rate constants were used. This led us to refine the model and predict a mechanism for release of chunks of the actin meshwork via severing by ADF/cofilin.

5. Explore the space of free parameters and perform a sensitivity analysis

The more parameters the model contains, the more likely several combinations of parameter values will fit the data. Best practice is to explore a range of reasonable parameter values to ensure the parameter set that best fits the data is indeed the best. In addition, exploring the parameter values around the values that best fit the data, also called sensitivity analysis, allows one to determine how robust the model is. If the model output changes dramatically with small changes in most of the parameters, it is likely that the parameter set is not biologically relevant. Indeed, one should expect that most biological systems are not too sensitive to small changes in most of their parameters, because evolution favors systems that are robust to small biological variabilities and environmental noise (Kitano 2007; Félix and Barkoulas 2015).

6. Validate the model

Once the model fits the data, it is important to verify that the assumptions of the model are still valid. For example, the continuum assumption that allows one to monitor concentrations instead of number of molecules might not be valid anymore if the number of molecules becomes too small during the simulations. Another example is that the simplifications made when developing theory (e.g., assuming one variable is much smaller than the others) should be checked in the final model. In the endocytic actin model, we assumed that the system was well mixed and the diffusion of unbound proteins was not reduced by the assembled actin meshwork. We verified *a posteriori* that the density of proteins at the endocytic site was still low enough not to reduce diffusion.

Another way to validate the model is to fit the model with only a subset of the information contained in the experimental data, then verify that the model also reproduces the unused data. For example, the cytokinetic ring model did not explicitly aim to reproduce the timing of ring formation, but simulations of the model that led to homogenous rings also assembled within the same timing as measured experimentally.

7. Test alternative hypotheses

A model is most powerful when it demonstrates not only that the modeled mechanisms can explain the observed experimental data, but also that competing alternative mechanisms cannot (Popper 1959; Howard 2014). If alternative mechanisms can also reproduce the data, one should explore both models and identify conditions where the output is different, and design experiments to test them (see below).

In the endocytic actin model, two alternative mechanisms for actin disassembly were possible: very fast depolymerization from the pointed ends (faster than rates measured *in vitro*), or standard depolymerization rates along with release of chunks of the meshwork after severing events. We identified that under the former hypothesis, a slight change in the assembly rate had dramatic effects, whereas the system was much more robust with the latter hypothesis.

8. Make testable predictions and test them experimentally

One of the most powerful features of mathematical models is their ability to make predictions, i.e., to identify conditions where the model outputs are unexpected or have never been explored experimentally before (Doudna et al. 2017). The most useful predictions are predictions that (a) can be easily tested experimentally and (b) have clearly distinguishable outcomes for the different competing hypotheses. Very good understanding of the experimental techniques used in one's field is required to evaluate whether the model's predictions can be experimentally tested and how easy the experiments are to perform. If one is not familiar with the experimental techniques in their field, one should not hesitate to consult with experimentalists. Ideally, once appropriate testable hypotheses have been found, one should aim to perform the experiments, potentially in a collaboration. Even if the experiments are not performed immediately, describing testable predictions along with the model makes the publication more impactful, and experimentalists will likely be inspired to perform them eventually.

The endocytic actin model made several predictions regarding the rate constants for actin dynamics and Arp2/3 complex branch formation in *Schizosaccharomyces pombe*, which were validated in several follow-up experiments (Ti and Pollard 2011; Arasada and Pollard 2011; Chen and Pollard

2013). The cytokinetic ring model predicted that reducing the severing rate of links between nodes has a dramatic effect on the homogeneity of the final ring, which was later confirmed using ADF/cofilin mutant with weak severing activity (Chen and Pollard 2011).

Even when experimental validation is not directly possible, the predictions might still be invaluable to discriminate between molecular mechanisms. For example, the endocytic actin model predicted that at most 8 barbed ends grow at a given time during endocytosis. This number is not directly measurable with methods currently available but it suggested that the force generated by actin polymerization alone (at most ~ 100 pN) is not sufficient to achieve membrane invagination (which requires more than 1000 pN) (Lacy et al. 2018), and inspired a search for new force production mechanisms that have been overlooked (Ma and Berro 2018).

9. Repeat

New experiments contradicting the model predictions, or new published data, are great opportunities to revisit the model's assumptions and restart the modeling process to challenge or refine our understanding of the underlying mechanisms.

Discussion

We presented our guidelines in a sequential order but regular interchanges and cycling between the different steps are likely during the model development. Developing a mathematical model forces one to dig more deeply in the literature of the field and learn more about what is known and what is unknown about the details of the system. These new collected pieces of information often lead one to revise some aspects of the approach, the hypotheses, or even the overarching question asked by the model. One may be tempted to increase the return on the time and resources invested in the model development and try to apply a variant of the model (or the modeling tools) to other biological questions. However, one should keep in mind that, like any other methodology in biology, the new model will be at best only as good (and useful) as the question asked.

Our guidelines highlight that developing useful models requires cross-disciplinary skills in order to be comfortable with the modeling tools and the biological system. Most importantly, cross-disciplinary skills allow one to translate the biological information into the mathematical formalism of a model, and conversely, to translate the outputs of the model into information that experimentalists can grasp and predictions that can be experimentally tested.

Mathematicians or physicists aiming to develop models in biology on their own or in collaboration with experimentalists should spend the time learning not only the biology of the

system they aim to model, but also the experimental methods classically used in that field. This will greatly help them determine what are good questions in the field, evaluate the accuracy of the published data, and get a sense of which experiments are realistically feasible to propose. Experimentalists who want to develop models or collaborate with theoreticians should learn the basics of mathematical modeling to get a better sense of what they should expect from the model. In both cases, learning the basics and knowing the mindset of each other's fields greatly eases communication and ensures a larger impact of the models and their predictions in the field (Drubin and Oster 2010).

Cross-disciplinary training should be encouraged at the undergraduate and graduate level (Bialek and Botstein 2004; Noble et al. 2016; Riveline and Kruse 2017), and scientists at any career stage who want to develop models should participate in cross-disciplinary workshops, networks and conferences, or teach their discipline to non-specialists. One should also take all opportunities to spend time in a lab from the other discipline, for example during a student exchange program, a postdoc or a faculty sabbatical, or by inviting people from other disciplines to spend some time in their lab. Being embedded in a different lab is an invaluable way to absorb their techniques, culture, language and standards, to identify the blind spots of each other's fields, and to learn how to communicate efficiently across disciplines.

Few researchers are equally comfortable with both experimental and modeling approaches and, for those who are, there are many low hanging fruits that can be picked and new research directions that they are uniquely positioned to explore. Attaining this fluency in both approaches requires time and practice. The path is not always easy because researchers in traditional disciplines too often consider cross-disciplinary projects as not “pure” enough for their discipline (Parsegian 1997), are uncomfortable with the intrinsically “dirty” nature of experimental data, or under-appreciate that fluency in multiple fields is more than the sum of its parts, and can lead to the creation of new branches of physics, mathematics, and biology (Phillips 2017). From experience, we think this effort is worth the trouble.

Acknowledgments The ideas developed in this paper have been inspired by my experience and the hurdles I have encountered developing models with experimentalist, mathematicians, and physicists, and transitioning from my initial field of applied mathematics to experimental biology while in Tom Pollard's lab. Although early in his career Tom had established himself as one of the most careful and insightful experimentalists, Tom very rapidly recognized the potential of mathematical modeling to study the cytoskeleton. Over the years, he has developed an exquisite intuition for mathematical modeling, and for working with researchers from diverse scientific backgrounds to build models that have been critical to advance our understanding of the biochemical, biophysical, and cellular properties of the actin cytoskeleton. I am extremely grateful Tom believed it was a good idea to encourage a mathematician to touch the pipettors in his lab.

The writing of this paper was catalyzed by discussions held at a Quantitative Cell Biology Network workshop (organized by Wallace Marshall and sponsored by NSF grant MCB-1411898), and with Enrique De La Cruz, without whom this paper would not exist. I thank Mike Lacy, Matt Akamatsu, Joe Howard, Shirin Bahmanyar, Rui Ma, Yuan Ren, Ronan Fernandez, and Ruchira Ray for their insights on the manuscript.

Funding information The research in the Berro lab is supported by National Institutes of Health/National Institute of General Medical Sciences Grant R01GM115636.

Compliance with ethical standards

Conflict of interest Julien Berro declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

References

- Andrews SS, Wren J (2017) Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. *Bioinformatics* 33:710–717. <https://doi.org/10.1093/bioinformatics/btw700>
- Arasada R, Pollard TD (2011) Distinct roles for F-BAR proteins Cdc15p and Bzz1p in actin polymerization at sites of endocytosis in fission yeast. *Curr Biol* 21:1450–1459. <https://doi.org/10.1016/j.cub.2011.07.046>
- Beltzner CC, Pollard TD (2008) Pathway of actin filament branch formation by Arp2/3 complex. *J Biol Chem* 283:7135–7144. <https://doi.org/10.1074/jbc.M705894200>
- Berro J, Lacy MM (2018) Quantitative biology of endocytosis. Morgan & Claypool, San Rafael
- Berro J, Pollard TD, Sirotkin V (2010a) What we learned from mathematical modeling of actin patch kinetics during endocytosis in fission yeast. 49th ASCB annual meeting. <https://www.ascb.org/wp-content/uploads/2015/12/2010ASCBFullProgram1.pdf>. Accessed 6 Nov
- Berro J, Sirotkin V, Pollard TD (2010b) Mathematical modeling of endocytic actin patch kinetics in fission yeast: disassembly requires release of actin filament fragments. *Mol Biol Cell* 21:2905–2915. <https://doi.org/10.1091/mbc.e10-06-0494>
- Bialek W, Botstein D (2004) Introductory science and mathematics education for 21st-century biologists. *Science* 303:788–790. <https://doi.org/10.1126/science.1095480>
- Blanchoin L, Pollard TD (1999) Mechanism of interaction of Acanthamoeba actophorin (ADF/Cofilin) with actin filaments. *J Biol Chem* 274:15538–15546. <https://doi.org/10.1074/jbc.274.22.15538>
- Blinov ML, Schaff JC, Vasilescu D et al (2017) Compartmental and spatial rule-based modeling with virtual cell. *Biophys J* 113:1365–1372. <https://doi.org/10.1016/j.bpj.2017.08.022>
- Boutillier P, Maasha M, Li X et al (2018) The kappa platform for rule-based modeling. *Bioinformatics* 34:i583–i592. <https://doi.org/10.1093/bioinformatics/bty272>
- Box GEP, Draper NR (1987) Empirical model-building and response surfaces, 1st edn. Wiley, New York
- Chang F, Drubin D, Nurse P (1997) cdc12p, a protein required for cytokinesis in fission yeast, is a component of the cell division ring and interacts with profilin. *J Cell Biol* 137:169–182. <https://doi.org/10.1083/jcb.137.1.169>

- Chang FS, Stefan CJ, Blumer KJ (2003) A WASp homolog powers actin polymerization-dependent motility of endosomes in vivo. *Curr Biol* 13:455–463. [https://doi.org/10.1016/S0960-9822\(03\)00131-3](https://doi.org/10.1016/S0960-9822(03)00131-3)
- Chen Q, Pollard TD (2011) Actin filament severing by cofilin is more important for assembly than constriction of the cytokinetic contractile ring. *J Cell Biol* 195:485–498. <https://doi.org/10.1083/jcb.201103067>
- Chen Q, Pollard TD (2013) Actin filament severing by cofilin dismantles actin patches and produces mother filaments for new patches. *Curr Biol* 23:1154–1162. <https://doi.org/10.1016/j.cub.2013.05.005>
- Doudna J, Bar-Ziv R, Elf J et al (2017) How will kinetics and thermodynamics inform our future efforts to understand and build biological systems? *Cell Syst* 4:144–146. <https://doi.org/10.1016/j.cels.2017.02.005>
- Drubin DG, Oster G (2010) Experimentalist meets theoretician: a tale of two scientific cultures. *Mol Biol Cell* 21:2099–2101. <https://doi.org/10.1091/mbc.E10-02-0143>
- Dyson F (2004) A meeting with Enrico Fermi. In: *Nature*. <https://www.nature.com/articles/427297a>. Accessed 9 Jul 2018
- Félix M-A, Barkoulas M (2015) Pervasive robustness in biological systems. *Nat Rev Genet* 16:483–496. <https://doi.org/10.1038/nrg3949>
- Flexner A (1939) The usefulness of useless knowledge. *Harpers Mag* 544–552. <https://harpers.org/archive/1939/10/the-usefulness-of-useless-knowledge/>. Accessed 6 Nov
- Fujiwara I, Vavylonis D, Pollard TD (2007) Polymerization kinetics of ADP- and ADP-Pi-actin determined by fluorescence microscopy. *Proc Natl Acad Sci* 104:8827–8832. <https://doi.org/10.1073/pnas.0702510104>
- Gachet Y, Hyams JS (2005) Endocytosis in fission yeast is spatially associated with the actin cytoskeleton during polarised cell growth and cytokinesis. *J Cell Sci* 118:4231–4242. <https://doi.org/10.1242/jcs.02530>
- Gilliam T, Jones T (1975) *Monty Python and the Holy Grail*. EMI Films. <https://www.imdb.com/title/tt0071853/>. Accessed 6 Nov
- Goldstein RE (2018) Point of view: are theoretical results ‘Results’? In: *eLife*. <https://elifesciences.org/articles/40018>. Accessed 21 Aug 2018
- Gostner R, Baldacci B, Morine MJ, Priami C (2014) Graphical modeling tools for systems biology. *ACM Comput Surv* 47:16:1–16:21. <https://doi.org/10.1145/2633461>
- Gunawardena J (2014) Models in biology: ‘accurate descriptions of our pathetic thinking’. *BMC Biol* 12:29. <https://doi.org/10.1186/1741-7007-12-29>
- Harris LA, Hogg JS, Tapia J-J et al (2016) BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics* 32:3366–3368. <https://doi.org/10.1093/bioinformatics/btw469>
- Hoops S, Sahle S, Gauges R et al (2006) COPASI—a Complex Pathway Simulator. *Bioinformatics* 22:3067–3074. <https://doi.org/10.1093/bioinformatics/btl485>
- Howard J (2014) Quantitative cell biology: the essential role of theory. *Mol Biol Cell* 25:3438–3440. <https://doi.org/10.1091/mbc.e14-02-0715>
- Hucka M, Finney A, Sauro HM et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531. <https://doi.org/10.1093/bioinformatics/btg015>
- Idrissi F-Z, Grötsch H, Fernández-Golbano IM et al (2008) Distinct actin/myosin-I structures associate with endocytic profiles at the plasma membrane. *J Cell Biol* 180:1219–1232. <https://doi.org/10.1083/jcb.200708060>
- Kaksonen M, Sun Y, Drubin DG (2003) A pathway for association of receptors, adaptors, and actin during endocytic internalization. *Cell* 115:475–487. [https://doi.org/10.1016/S0092-8674\(03\)00883-3](https://doi.org/10.1016/S0092-8674(03)00883-3)
- Kaksonen M, Toret CP, Drubin DG (2005) A modular design for the clathrin- and actin-mediated endocytosis machinery. *Cell* 123:305–320. <https://doi.org/10.1016/j.cell.2005.09.024>
- Kaksonen M, Toret CP, Drubin DG (2006) Harnessing actin dynamics for clathrin-mediated endocytosis. *Nat Rev Mol Cell Biol* 7:404–414. <https://doi.org/10.1038/nrm1940>
- Kim K, Galletta BJ, Schmidt KO et al (2006) Actin-based motility during endocytosis in budding yeast. *Mol Biol Cell* 17:1354–1363. <https://doi.org/10.1091/mbc.e05-10-0925>
- Kitano H (2007) Towards a theory of biological robustness. *Mol Syst Biol* 3:137. <https://doi.org/10.1038/msb4100179>
- Kovar DR, Kuhn JR, Tichy AL, Pollard TD (2003) The fission yeast cytokinesis formin Cdc12p is a barbed end actin filament capping protein gated by profilin. *J Cell Biol* 161:875–887. <https://doi.org/10.1083/jcb.200211078>
- Kovar DR, Wu J-Q, Pollard TD (2005) Profilin-mediated competition between capping protein and formin Cdc12p during cytokinesis in fission yeast. *Mol Biol Cell* 16:2313–2324. <https://doi.org/10.1091/mbc.e04-09-0781>
- Kuhn JR, Pollard TD (2007) Single molecule kinetic analysis of actin filament capping polyphosphoinositides do not dissociate capping proteins. *J Biol Chem* 282:28014–28024. <https://doi.org/10.1074/jbc.M705287200>
- Lacy MM, Ma R, Ravindra NG, Berro J (2018) Molecular mechanisms of force production in clathrin-mediated endocytosis. *FEBS Lett*. <https://doi.org/10.1002/1873-3468.13192>
- Lopez CF, Muhlich JL, Bachman JA, Sorger PK (2013) Programming biological models in Python using PySB. *Mol Syst Biol* 9:646. <https://doi.org/10.1038/msb.2013.1>
- Lord M, Laves E, Pollard TD (2005) Cytokinesis depends on the motor domains of myosin-II in fission yeast but not in budding yeast. *Mol Biol Cell* 16:5346–5355. <https://doi.org/10.1091/mbc.e05-07-0601>
- Ma R, Berro J (2018) Structural organization and energy storage in crosslinked actin assemblies. *PLoS Comput Biol* 14:e1006150. <https://doi.org/10.1371/journal.pcbi.1006150>
- Marshall WF (2017) Introduction to quantitative cell biology. Morgan & Claypool Life Sciences, San Rafael
- Matsumura F (2005) Regulation of myosin II during cytokinesis in higher eukaryotes. *Trends Cell Biol* 15:371–377. <https://doi.org/10.1016/j.tcb.2005.05.004>
- Mayer J, Khairy K, Howard J (2010) Drawing an elephant with four complex parameters. *Am J Phys* 78:648–649. <https://doi.org/10.1119/1.3254017>
- Michaelis L, Menten ML (1913) Die Kinetik der Invertinwirkung. *Biochem Z* 49:333–369
- Möbius W, Laan L (2015) Physical and mathematical modeling in experimental papers. *Cell* 163:1577–1583. <https://doi.org/10.1016/j.cell.2015.12.006>
- Mogilner A, Wollman R, Marshall WF (2006) Quantitative modeling in cell biology: what is it good for? *Dev Cell* 11:279–287. <https://doi.org/10.1016/j.devcel.2006.08.004>
- Mogilner A, Allard J, Wollman R (2012) Cell polarity: quantitative modeling as a tool in cell biology. *Science* 336:175–179. <https://doi.org/10.1126/science.1216380>
- Motegi F, Nakano K, Mabuchi I (2000) Molecular mechanism of myosin-II assembly at the division site in *Schizosaccharomyces pombe*. *J Cell Sci* 113:1813–1825
- Naqvi NI, Wong KCY, Tang X, Balasubramanian MK (2000) Type II myosin regulatory light chain relieves auto-inhibition of myosin-heavy-chain function. *Nat Cell Biol* 2:855–858. <https://doi.org/10.1038/35041107>
- Noble DB, Mochrie SGJ, O’Hern CS et al (2016) Promoting convergence: the integrated graduate program in physical and engineering biology at Yale University, a new model for graduate education. *Biochem Mol Biol Educ* 44:537–549. <https://doi.org/10.1002/bmb.20977>
- O’Shaughnessy B, Pollard TD (2016) Mechanistic biological modeling thrives. *Science* 351:234–235. <https://doi.org/10.1126/science.351.6270.234-c>

- Parsegian VA (1997) Harness the hubris: useful things physicists could do in biology. *Phys Today* 50:23–27. <https://doi.org/10.1063/1.881805>
- Phillips R (2015) Theory in biology: figure 1 or figure 7? *Trends Cell Biol* 25:723–729. <https://doi.org/10.1016/j.tcb.2015.10.007>
- Phillips R (2017) Musings on mechanism: quest for a quark theory of proteins? *FASEB J* 31:4207–4215. <https://doi.org/10.1096/fj.201700594>
- Pollard TD (2010) A guide to simple and informative binding assays. *Mol Biol Cell* 21:4061–4067. <https://doi.org/10.1091/mbc.e10-08-0683>
- Pollard TD (2013) No question about exciting questions in cell biology. *PLoS Biol* 11:e1001734. <https://doi.org/10.1371/journal.pbio.1001734>
- Pollard TD, De La Cruz EM (2013) Take advantage of time in your experiments: a guide to simple, informative kinetics assays. *Mol Biol Cell* 24:1103–1110. <https://doi.org/10.1091/mbc.e13-01-0030>
- Popper KR (1959) *The logic of scientific discovery*. Hutchinson & Co., Ltd., London, p 480
- Riveline D, Kruse K (2017) Interface between physics and biology: training a new generation of creative bilingual scientists. *Trends Cell Biol* 27:541–543. <https://doi.org/10.1016/j.tcb.2017.05.002>
- Roland J, Berro J, Michelot A et al (2008) Stochastic severing of actin filaments by actin depolymerizing factor/cofilin controls the emergence of a steady dynamical regime. *Biophys J* 94:2082–2094. <https://doi.org/10.1529/biophysj.107.121988>
- Rothman JE (2018) Jim's view: "Some Thoughts for Young Scientists.". *FEBS Lett* 592:461–462. <https://doi.org/10.1002/1873-3468.12960>
- Sirotkin V, Beltzner CC, Marchand J-B, Pollard TD (2005) Interactions of WASp, myosin-I, and verprolin with Arp2/3 complex during actin patch assembly in fission yeast. *J Cell Biol* 170:637–648. <https://doi.org/10.1083/jcb.200502053>
- Sirotkin V, Berro J, Macmillan K et al (2010) Quantitative analysis of the mechanism of endocytic actin patch assembly and disassembly in fission yeast. *Mol Biol Cell* 21:2894–2904. <https://doi.org/10.1091/mbc.e10-02-0157>
- Slepchenko BM, Schaff JC, Carson JH, Loew LM (2002) Computational cell biology: spatiotemporal simulation of cellular events. *Annu Rev Biophys Biomol Struct* 31:423–441. <https://doi.org/10.1146/annurev.biophys.31.101101.140930>
- Thomas M, Schwartz R (2017) Quantitative computational models of molecular self-assembly in systems biology. *Phys Biol* 14:035003. <https://doi.org/10.1088/1478-3975/aa6cdc>
- Ti S-C, Pollard TD (2011) Purification of actin from fission yeast *Schizosaccharomyces pombe* and characterization of functional differences from muscle actin. *J Biol Chem* 286:5784–5792. <https://doi.org/10.1074/jbc.M110.199794>
- Vavylonis D, Wu J-Q, Hao S et al (2008) Assembly mechanism of the contractile ring for cytokinesis by fission yeast. *Science* 319:97–100. <https://doi.org/10.1126/science.1151086>
- Wick SM, Kane CM (2011) Cell biology education: where's the math? *Mol Biol Cell* 22:716–716. <https://doi.org/10.1091/mbc.e11-01-0029>
- Wu J-Q, Kuhn JR, Kovar DR, Pollard TD (2003) Spatial and temporal pathway for assembly and constriction of the contractile ring in fission yeast cytokinesis. *Dev Cell* 5:723–734. [https://doi.org/10.1016/S1534-5807\(03\)00324-1](https://doi.org/10.1016/S1534-5807(03)00324-1)
- Wu J-Q, Sirotkin V, Kovar DR et al (2006) Assembly of the cytokinetic contractile ring from a broad band of nodes in fission yeast. *J Cell Biol* 174:391–402. <https://doi.org/10.1083/jcb.200602032>