Title
How and why to build a mathematical model: A case study using prion aggregation

Permalink
https://escholarship.org/uc/item/7vh8q7mm

Journal
Journal of Biological Chemistry, 295(15)

ISSN
0021-9258

Authors
Banwarth-Kuhn, Mikahl
Sindi, Suzanne

Publication Date
2020-04-01

DOI
10.1074/jbc.rev119.009851

Peer reviewed

How and why to build a mathematical model: A case study using prion aggregation

Mikahl Banwarth-Kuhn and Suzanne Sindi[1]

From the Department of Applied Mathematics, School of Natural Sciences, University of California, Merced, California 95343

Edited by Enrique M. De La Cruz

Biological systems are inherently complex, and the increasing level of detail with which we are able to experimentally probe such systems continually reveals new complexity. Fortunately, mathematical models are uniquely positioned to provide a tool suitable for rigorous analysis, hypothesis generation, and connecting results from isolated *in vitro* experiments with results from *in vivo* and whole-organism studies. However, developing useful mathematical models is challenging because of the often different domains of knowledge required in both math and biology. In this work, we endeavor to provide a useful guide for researchers interested in incorporating mathematical modeling into their scientific process. We advocate for the use of conceptual diagrams as a starting place to anchor researchers from both domains. These diagrams are useful for simplifying the biological process in question and distinguishing the essential components. Not only do they serve as the basis for developing a variety of mathematical models, but they ensure that any mathematical formulation of the biological system is led primarily by scientific questions. We provide a specific example of this process from our own work in studying prion aggregation to show the power of mathematical models to synergistically interact with experiments and push forward biological understanding. Choosing the most suitable model also depends on many different factors, and we consider how to make these choices based on different scales of biological organization and available data. We close by discussing the many opportunities that abound for both experimentalists and modelers to take advantage of collaborative work in this field.

How living organisms function and biological systems work remain some of the most profound mysteries in our world. Many areas of biological inquiry aim to explain how complex behaviors emerge from simple rules of interaction between fundamental components. Two examples of this type of question are "How do collections of different cell types communicate to develop into complex structures such as an entire organ or embryo?" and "How do microscopic protein aggregation dynamics and macroscopic cell behaviors within a growing tissue interact to propagate or clear neurodegenerative diseases such as Alzheimer's?" Answering these questions is challenging, and requires combining approaches from different domains. In particular, one of the most successful combinations of expertise is that of experimental and biological sciences with the mathematical and computational sciences. The use of mathematical models was proven to be fundamental toward advancing physics in the 20th century, and many are projecting mathematics to play a similar role in advancing biological discovery in the 21st century (1, 3).

Mathematical and computational models have already begun to play an increasingly large role in the advancement of biological and biochemical research because they make it possible to quantitatively bridge the gap between data gathering and mechanism testing by providing a set of analytical and numerical tools. The level at which we are able to collect data ranges across many different scales, including single-cell analysis, molecular interactions, tumor growth, and epidemiological statistics. In most cases, we aim to understand how the processes and mechanisms we track at microscopic scales lead to emergent patterns of disease and other behaviors observed at the macroscopic scale of entire colonies, tissues, or populations. The aim of combined experimental investigation and mathematical and computational modeling in the biological sciences is to develop a tool set that can combine information at multiple scales and elucidate the underlying mechanisms that drive an observed phenomenon (Fig. 1).

Experimental studies are rooted in observation to draw conclusions about how a biological system functions, and they rely on experimental manipulations to clarify which components are the most important. However, it can be challenging to make the required observations and manipulations without affecting the system in unintentional ways. Mathematical and computational models use observation and manipulations in the same way as experiments, but they can avoid many of the most challenging experimental difficulties. In addition, such models often permit experiments that are currently not feasible in the physical system. In the best cases, mathematical models complement experimental studies by providing new insight on the most crucial interactions within the system. Although mathematical descriptions of biological phenomena cannot replace biological experiments or provide biological evidence that a particular mechanisms is at work, they can demonstrate
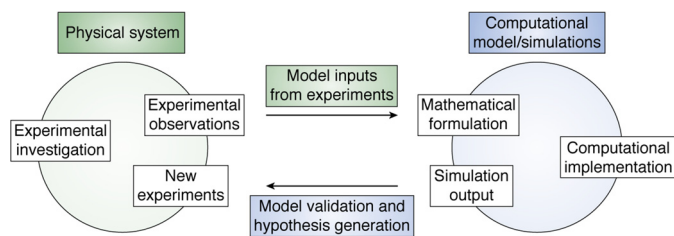
This article contains supporting information.
[1] To whom correspondence should be addressed. Tel.: 209-228-4224; E-mail: ssindi@ucmerced.edu.

Figure 1. Interplay between experiments and mathematical models. Mathematical models can be used to test hypotheses, probe changes in parameters, generate predictions, and design new experiments.

whether or not a proposed mechanism is sufficient to produce an observed phenomenon.

Before access to powerful computers was so common, it was possible to create and analyze a mathematical model by hand. However, using a handwritten equation to study realistic biological applications, such as interactions between meaningful numbers of cells or many different biochemical species, becomes impractical due to the large number of equations and variables. Once computers became available, code could be written to automatically construct and solve equations repeatedly over multiple time steps. Thus, computational implementation of models is an important feature that brings together different kinds of data across a range of length scales. Moreover, reasonable knowledge of computational methods provides a new way for researchers to investigate interactions between different system elements on a very large scale. Even in the case of biological mechanisms that are well-understood, mathematical and computational modeling makes it possible to explore the consequences of manipulating various parameters, which in the case of brain tumors and other cancers has become a good resource for simulating different treatment protocols before applying them in practice (4–11).

The purpose of this article is to empower researchers in the biological and experimental sciences to develop their own models as well as to lower the barrier between these fields with the mathematical sciences. Indeed, there are significant challenges in interdisciplinary collaborations. Both parties—the mathematical/computational and experimental/biological—may feel that they do not have sufficient expertise in the opposite domain to begin a fruitful conversation. To paraphrase James A. Yorke—an applied mathematician credited with coining the term "chaos theory," who has a long track record of productive interdisciplinary work—"interdisciplinary research requires being comfortable asking 'naive' questions." The purpose of this review is to help communicate some of the language surrounding mathematical modeling in a way that will facilitate productive interactions between scientists and to demonstrate that for both sides, trekking into the great unknown is not only intellectually rewarding, but offers the potential to introduce significant advances in both fields.

We structure this Review around illustrating a pipeline for mathematical model building in an interdisciplinary setting that has been productive for us in the past. First, we give an overview of mathematical modeling in general and provide a three-step process we have found useful in model development. We then illustrate this process using a canonical enzyme substrate reaction as a guide. In Section 3 we discuss a more com-

plex biochemical system, the aggregation and fragmentation dynamics of prion aggregates. Although this system consists of a (theoretically) infinite number of interacting biochemical species, the same three-step process can be used to develop a mathematical model. We then demonstrate how this model provides a toolbox for answering scientific questions about prion disease. Finally, we close with a discussion of considerations for developing more complex models at different scales and provide resources for readers interested in building their own models. In particular, all of the code for the models we develop in this Review is available in the supporting information as an iPython Notebook. We encourage readers to explore these models on their own. (We also note that there are many books and resources delving into this material in greater detail than possible in this Review. We encourage those interested in learning more about mathematical modeling to explore other reviews with examples from different applications written to guide researchers in building models in biology (12–15) or textbooks in mathematical biology (16–22).)

What does it mean to model something?

Those new to mathematical modeling may be curious what it actually means to construct a model. We emphasize that to most effectively drive scientific discovery, a model cannot exist in a vacuum but must be intimately involved in the experimental process. In Fig. 1, we demonstrate some interactions between an experimental investigation and a mathematical and computational model. We encourage experimental researchers to think of a mathematical model itself as an experimental tool. The mathematical model can be used to make a prediction about what will happen under certain conditions (*i.e.* different initial conditions and/or parameters) and observe whether or not the predicted outcome occurs. The mathematical model can be used to study how sensitive one output of interest is to increasing or decreasing the amount of other factors. The mathematical model can also be used to aid in the design of new experiments. However, this process is not one-sided. Developing a useful mathematical model relies on having a solid understanding of the system under study. Unknown parameters may be "fit" by comparing model output with data, and the model itself can be "validated" by predicting an unexpected experimental outcome.

We emphasize three steps in the design of a mathematical model. The first step in building a mathematical and computational model is to formulate a diagram that specifies the key players (state variables) and describes all possible ways these variables might interact with each other. For example, one of the most basic enzymatic reactions, first proposed by Michaelis and Menten (23), involves a substrate S reacting with an enzyme E to form an enzyme substrate complex $E:S$, which is converted into a product P and the enzyme. The diagram for this system is given in Fig. 2 (*left*). Note that this diagram is by nature conceptual; we are typically specifying only the interactions and not the exact quantitative nature of those interactions. We emphasize that the creation of this type of diagram is particularly important in interdisciplinary work because its visual nature makes it accessible to all researchers and provides a common ground for moving forward.
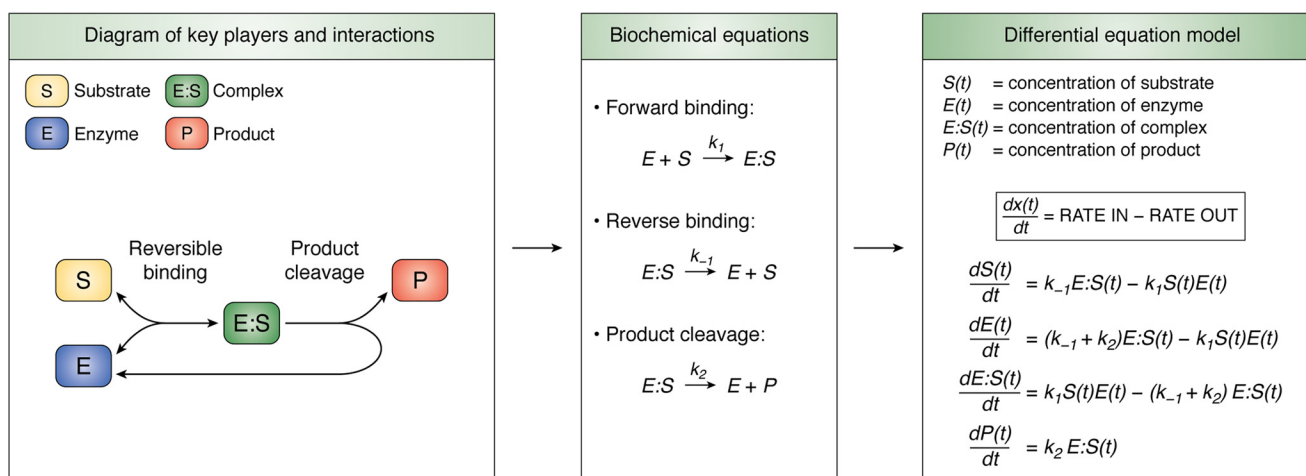
Figure 2. Building a model of enzyme substrate kinetics. *Left*, a conceptual diagram illustrating the key biochemical species important in the system along with their interactions. *Middle*, explicit description of the biochemical reactions represented from the diagram. *Right*, a mathematical model, a system of ordinary differential equations, describing the rate of change of each biochemical species.

The second step in building a mathematical and computational model is to explicitly write out the quantitative form of all interactions. This is shown for our example system in Fig. 2 (*middle*). In our experience studying biological systems at the molecular scale, biochemical equations provide a natural way to list these interactions. (For a discussion of how this step is handled at different biological scales, see Section 4.) It is also at this step that decisions about stoichiometry and the form of interactions are clarified (it is possible that complex interactions are simplified and multistep reactions determined) and rates (more so as symbols than numerical values) are assigned to specific reactions. Most importantly, this step requires gathering knowledge from experiments and experts in the field to incorporate what is currently understood about the system of interest.
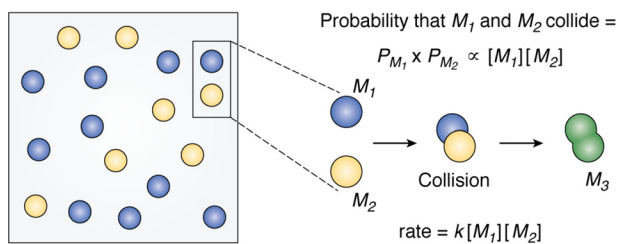
In our example, we see that the enzyme (E) and substrate (S) form a complex ($E{:}S$) at rate k_1 but that this complex itself is reversible at rate k_{-1}. This complex may also, at rate k_2, result in the creation of a product (P) and a return of enzyme in the complex to the free pool. (In this example, the system as defined does not include synthesis or degradation, so mass is preserved. This simplification is reflected in our biochemical equations because there is no synthesis or degradation and we only show three reaction rates.)

The third step in constructing a mathematical model is to convert the explicit quantitative interactions into a mathematical framework. In general, there are many questions to ask before completing this step. Are the biochemical species considered to be concentrations or numbers? Is the time scale discrete or continuous? Is the system deterministic or stochastic? The resulting mathematical formulation must include a clear explanation of which system components are being modeled, how each component is represented (*i.e.* integer, real number, Boolean, etc.), and why the choice of representation is appropriate for each component. (For simplicity and because of their popularity, in this paper, we will consider a deterministic model of concentrations. We discuss other possibilities in Section 4.)

How can we convert our conceptual model into a mathematical framework? For differential equation models that aim to

capture molecular dynamics, the standard approach is to use The law of mass action (see Fig. 3). This law states that the rate of a reaction is proportional to the product of the concentrations of the reactants. It is this rate, the product of the reaction rate and the product of the reactants, which determines the instantaneous change in concentration of all biochemical species. In general, the mass action assumption provides a tool for setting up a general mathematical framework by defining the relationship between state variables (*i.e.* proportional) in terms of measurable quantities (*i.e.* concentration of the reactants). In our example, the biochemical reactions lead to the well-studied system of differential equations depicted in Fig. 2 (*right*). (We note that although the law of mass action was developed for molecular processes, it has been used as a model for interactions at larger scales and indeed is the foundation for many mathematical models in ecology and epidemiology (19–22).)

After the model is built, the real work begins! The model may be analyzed theoretically (by considering the form of the differential equations directly) or numerically (by considering the time-varying changes in the quantities of interest). The first task is typically to investigate whether the model behaves in a manner consistent with the knowledge of the system. For example, the enzyme-substrate system we considered has been studied extensively. Over time, all of the substrate will be converted to product, and the shape of the product curve can be used to determine relationships between the parameters. Once the model is considered to be consistent with what is known about a system, the model is then probed to reveal aspects that would be difficult or impossible to uncover experimentally. For example, to test the sensitivity of the system to changing initial conditions, a researcher could set up thousands of experiments, each with different conditions. Alternatively, the mathematical model could be more easily probed for parameter sensitivity. In addition, it is often only possible to experimentally observe a subset of the state variables of interest (*i.e.* usually only the product concentration $P(t)$ can be visualized). However, the mathematical model can reveal the "hidden" concentrations of all other variables too. Because we suspect most readers are familiar with this simple system, in the next section we explore

Probability that M_1 and M_2 collide =

$$P_{M_1} \times P_{M_2} \propto [M_1][M_2]$$

M_1

M_2

Collision

M_3

rate = $k[M_1][M_2]$

Figure 3. The law of mass action. The relationship of the rate of a reaction to the concentrations of the chemical species that are involved in the reaction is given by the law of mass action. This law states that the rate of a reaction is proportional to the product of the concentrations of the reactants.

developing a mathematical model in a more complicated setting and demonstrate how analytical and numerical methods are used to study the model.

Case study: Protein aggregation kinetics

In the previous section, we went over the basics of building a mathematical model through 1) creating a diagram of the critical players (state variables) and their interactions; 2) enumerating the complete set of interactions, typically as a set of biochemical equations; and 3) translating these interactions into a mathematical model, typically through the law of mass action. Because we suspect many readers are familiar with Michaelis–Menten kinetics, we now consider the same modeling guidelines to construct a mathematical model used in the study of prion aggregate dynamics.

Prions are associated with a number of progressive, incurable, and fatal neurological diseases in mammals (24, 25). Remarkably, these diseases may be either spontaneous, genetic, or acquired (26–28). These different modes of transmission and extremely long incubation times made it challenging for researchers to determine the infectious agent (29–31). However, today the scientific consensus is that they are caused by a proteinaceous infectious agent, which is where the term prion comes from (32–35) (see Ref. 36 for a detailed history on the study of prion disease). Whereas the true biochemical and biophysical processes of prion aggregate dynamics are highly complicated and may differ depending on strain and infected tissue, mathematical modeling of idealized biochemical processes has been extremely effective at uncovering critical steps in the pathway (37–41). Before describing a mathematical framework for modeling the dynamics of prion proteins, we consider what is known about the biological system. We will then go through the underlying development of a diagram identifying the key players and their interactions. Then we will discuss a mathematical model based on the biochemistry and discuss what it enables us to learn about the biology.

All mammalian prion diseases are a consequence of the same protein, PrP (prion-domain protein). In these diseases, a misfolded form of the PrP (PrP-C) is introduced to the host and, rather than being eliminated or cleared by cellular quality control mechanisms, this misfolded (prion) form persists and induces other normally folded proteins in the host to fold into their same conformation (42). This conversion process is thought to happen through associations between normally folded protein and aggregates of the prion form. The size of these prion aggregates thus increases by incorporating the newly misfolded protein (polymerization). The size of these aggregates may also decrease either through depolymerization (which removes a single monomer) or fragmentation (which then can amplify the total number of aggregates). These aggregates then spread through the mammalian host. Remarkably, prion diseases are not only confined to mammals. The single-cell fungus *Saccharomyces cerevisiae* has several harmless phenotypes that we now know to be caused by prion proteins (34). Moreover, certain human neurodegenerative diseases (Parkinsons's, Huntington's, etc.) propagate by a similar aggregation process, as was demonstrated by a recent study of Alzheimer's disease (43). Finally, although prion aggregates are extremely thermodynamically stable, the appearance of prion disease is thought to be nucleation-dependent (44). In other words, the aggregates can be thought of as a protein phase transition that necessitates the formation of a stable nucleus consisting of some number, n_0, of misfolded protein monomers (45). (For the $[PSI^+]$ prion in *S. cerevisiae*, the nucleus is thought to consist of five misfolded monomers (37, 40).) Aggregates below n_0 in size are highly unstable and are thought to rapidly resolve into monomers (35, 40, 44, 46).

Setting the stage for a mathematical model

After having reviewed the biological background on prion aggregation, the next step is to determine the key biochemical players and how they will interact. This step is particularly important, as it sets the stage for how complex our model will ultimately become. In the case of prion dynamics, we know that there are a host of biochemical factors that are important for disease propagation. But it is usually best to begin with the simplest set of assumptions and add complexity only when needed. (For a discussion of how to further develop our mathematical model to include more complex interactions, see Section 4.) In what follows, we lead the reader through the development of the nucleated polymerization model (NPM).[2] This model has become the standard choice for modeling prion aggregation dynamics (44, 46). First, based on the biology we know, we need to consider two classes of biochemical species: soluble protein monomers and aggregates of the prion (misfolded) form of the protein. Because we know that aggregates consist of groups of misfolded proteins, we will consider aggregates of every discrete size larger than the critical nucleus size n_0.

The next question we ask is: How will these species interact? We know that there is some process by which normal protein (a soluble protein monomer) is converted to the prion form and incorporated into an existing aggregate and some other process by which aggregates are fragmented. So we will need to specify these interactions. In addition, following the common critical nucleus size assumption, any aggregate below the nucleus size n_0 in our model will be considered not stable and quickly resolved into healthy monomer. As we said before, once the species and interactions are identified, it is advisable to codify these properties in a picture or diagram. In addition to being a useful tool for a mathematical modeler on its own, the visual medium of pictures facilitates communication between

---

[2] The abbreviation used is: NPM, nucleated polymerization model.
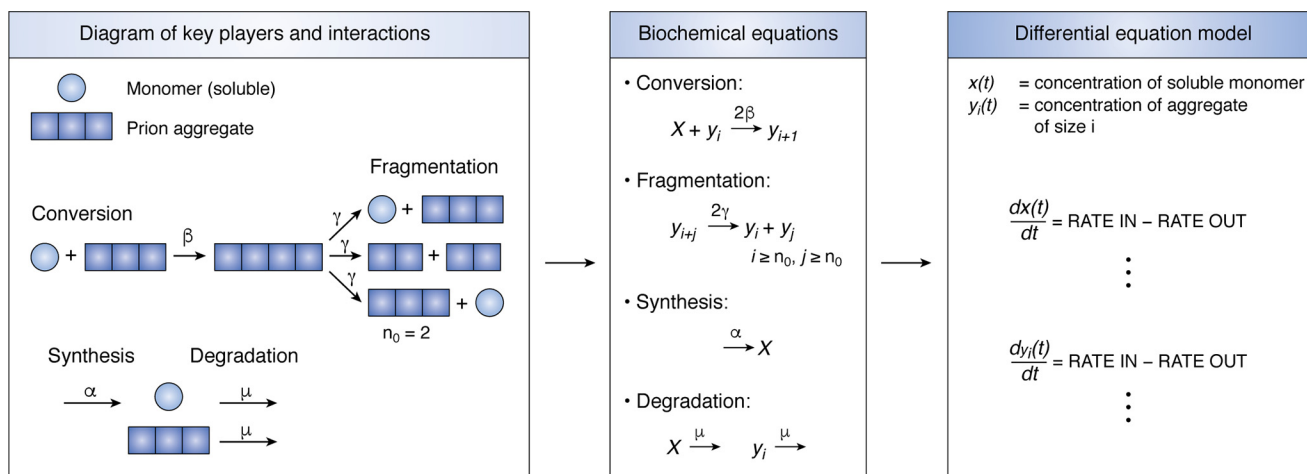
Figure 4. Diagram to differential equations. *Left*, a diagram depicting the key players (*circle*, monomer; *blocks*, aggregate) and their interactions. *Middle*, the set of biochemical equations that govern the interactions between the key biochemical players monomer X and aggregates of size $i = y_i$. (Note that because of the assumption of a minimum stable nucleus size of n_0, our form of the fragmentation equation depicted is correct only if $i \geq n_0, j \geq n_0$). *Right*, differential equation model schematic depicting the temporal evolution of the concentration of monomer ($x(t)$) and aggregates of each size i ($y_i(t)$). See Section 3 for more information.

researchers in different domains. Agreeing on a diagram of interactions is an important first step in mathematical modeling.

We show a diagram depicting our prion interactions in Fig. 4 (*left*). We consider two types of biochemical species: soluble (monomer) and aggregates. We have also introduced the idea that the number of blocks in an aggregate indicates its size (*i.e.* conversion shows an aggregate of size 3 becoming an aggregate of size 4). We have also diagramed the basic nature of the conversion and fragmentation reactions. Notice that in drawing our fragmentation equation, we had to make a decision about what happens when fragmentation creates an aggregate with a size below that of the nucleus ($n_0 = 2$). Whereas one choice could be that those proteins completely degrade, we have instead decided to have these proteins return to the soluble monomer state. (This is consistent with the NPM (44, 46).) Our diagram also depicts two additional reactions: monomer is created (rate α), and both monomer and aggregates are degraded (rate μ). The inclusion of such processes depends on the system under consideration. (Because the original NPM was developed to depict *in vivo* prion kinetics, we assume the protein synthesis and degradation machinery to be functioning. Of course, aggregation models of *in vitro* aggregation may elect to consider different processes, as in Ref. 47.)

Once the diagram is defined, the next step is to write out the steps in the diagram explicitly as a series of biochemical equations (see Fig. 4 (*middle*)). The *left-hand side* of the biochemical equations gives the reactants, what is needed or consumed for the reaction to be carried out, and the *right-hand side* shows the resulting products. The reaction rate designates the speed at which the reaction takes place. The units of the reaction rate typically depend on the order (number of reactants) of the reaction itself. (For example, in a first-order reaction (only a single reactant), the reaction rate has units of (time)$^{-1}$.)

It is during this step that the details of our interactions are refined and formalized. In the case of our conversion equation, we might consider the size of an aggregate to impact the rate of conversion. For example, an aggregate of size 10 could more

easily convert than an aggregate of size 5 if every part of its surface was capable of templating. In the case of the NPM, prion aggregates are considered to be amyloid—an ordered linear structure of misfolded protein. As such, conversion of monomer is only considered possible at the ends of the amyloid. As such, no matter the size of the aggregate, it has only two sites (either end) where conversion can occur. This is reflected in the top biochemical equation in Fig. 4 (*middle*).

A similar decision is necessary for the fragmentation reaction. It is possible that not all sites in a linear aggregate are equally frangible. (Indeed, in some *in vitro* experiments, the best fitting mathematical model was one in which the middle of amyloid fibers was more likely to fragment than the ends (48).) The assumption taken by the NPM is that each junction between monomers is equally likely. In this case, if each junction in an aggregate fragments at the same rate γ, then given an aggregate of size ($i + j$), there are two fragmentation sites that would result in two aggregates with distinct sizes (i and j). (Note that if $i = j$, then although there is only one site that could create these aggregates, the stoichiometry remains the same, because we produce two aggregates of the same size.) Similar biochemical reactions govern the processes of synthesis (α) and degradation (μ). (Note that in the original NPM, distinct degradation rates were considered for monomers and aggregates. However, for simplicity, we consider both processes to occur at the same rate μ.)

Development and analysis of the nucleated polymerization model: A mathematical model of prion aggregation

With the first two steps complete, the next phase is to design and analyze an appropriate mathematical model. Before deciding on a mathematical framework for our model, we need to decide the types of scientific questions we want to answer. In this exercise, we will consider two questions of interest:

- Can we determine what causes prion aggregates to be cleared by manipulating biochemical rates?

- Why is there a long incubation time for many prion diseases?

Because these questions are based on the temporal evolution of prion aggregates, we will chose a differential equation model that tracks the changes in concentration of each biochemical species (see Fig. 4 (*right*)). In other words, we define $x(t)$ to be the concentration of healthy monomer and $y_i(t)$ to be the concentration of an aggregate of size i and track their evolving concentrations in time. We use the law of mass action to convert our biochemical equations into differential equations. As mentioned above, this law states that the instantaneous rate of a reaction is a product of a reaction rate and the product of the concentration of all reactants at that time. Let's look at an example. Suppose the current concentrations of the monomer x and aggregates y_i are known. What is the instantaneous rate of aggregate conversion of monomer by an aggregate of size i?

According to the law of mass action and the *top* biochemical equation in Fig. 4 (*middle*), this rate is given by 2β times the concentration of monomer x times the concentration aggregates of size i, y_i,

Rate of conversion of monomer by aggregate of size

$$i := 2\beta x y_i. \tag{Eq. 1}$$

We could repeat the steps for each possible reaction in our system. (Note that because our system considers all aggregate lengths larger than n_0, our system formally has infinitely many reactions to consider! We will soon see how this is no obstacle for mathematical modeling.) We then combine these reaction rates into a differential equation for each biochemical species. The differential equations are themselves simply the sum of reactions that create the given species (rate in) minus the reactions that consume the species (rate out). Let's consider the differential equation for the monomer concentration $x(t)$,

$$\frac{dx(t)}{dt} = \text{Rate In} - \text{Rate Out}. \tag{Eq. 2}$$

Let's first consider the "Rate Out" part of the equation. There are two types of reactions that consume monomer. First, monomer is being degraded at rate μ times the current concentration of $x(t)$. Second, monomer is consumed during conversion by every possible aggregate size. We have already calculated the rate of conversion by an aggregate of size i in Equation 1. In calculating this bulk rate, we need to sum the rate of conversion over all possible aggregate sizes,

$$\text{Rate Out} = \mu x(t) + 2\beta x(t) y_{n_0}(t) + 2\beta x(t) y_{n_0+1}(t) + \ldots$$

$$= \mu x(t) + 2\beta x(t) \sum_{i \geq n_0} y_i(t). \tag{Eq. 3}$$

Now we move on to the "rate in" portion of the differential equation for $x(t)$. As for the "rate out" process, there are two types of reactions that create monomer. First, monomer is synthesized at rate α. Second, monomer is created when aggregates are fragmented in such a way that one piece of the recently fragmented aggregate is below the minimum stable size.

To construct the "rate in" portion of the differential equation, we consider an aggregate of exactly the minimum stable size n_0. This aggregate has $n_0 - 1$ fragmentation sites, and every possible fragmentation event will result in the creation of two aggregates with size less than n_0. Due to our assumption that any aggregate below the nucleus size n_0 is not stable, each of the fragmented pieces will be resolved into n_0 monomers. Thus, the rate of monomer creation by fragmentation of aggregates of size n_0 is given by the following,

Rate of monomer creation by fragmenting an aggregate of size

$$n_0 := \gamma n_0 (n_0 - 1) y_{n_0}(t). \tag{Eq. 4}$$

For an aggregate of size $(n_0 + 1)$ there are n_0 fragmentation junctions. Two of them will create an aggregate of size n_0 and recover a single monomer, and the remaining $(n_0 - 2)$ junctions would result in all $(n_0 + 1)$ proteins in the aggregate returning to the monomer state. (For example, suppose $n_0 = 3$ and an aggregate of size 4 is going to be fragmented. There are three fragmentation sites, each of which is equally likely to be chosen. If the first or third fragmentation site is chosen, the resulting pieces will be size 1 and size 3. Because $n_0 = 2$, the size 1 piece will return to the monomer state, and the size 3 piece will remain an aggregate. If the second fragmentation site is chosen, the resulting pieces will both be 2, which is smaller than n_0, and the result will be 4 monomers.) Thus, the rate of monomer creation by fragmentation is given by the following,

Rate of monomer creation by fragmenting an aggregate of size

$$n_0 + 1 := \gamma(2 + (n_0 - 2)(n_0 + 1)) y_{n_0 + 1}(t)$$

$$= \gamma(2 + n_0^2 - 2n_0 + n_0 - 2) y_{n_0 + 1}(t)$$

$$= \gamma(n_0^2 - n_0) y_{n_0 + 1}(t)$$

$$= \gamma n_0 (n_0 - 1) y_{n_0 + 1}(t). \tag{Eq. 5}$$

Note that this has the same coefficient $\gamma n_0(n_0 - 1)$ as before. In fact, this pattern holds for all aggregate sizes but requires slightly different reasoning when the aggregate size exceeds $2n_0$. In this case, at most, one of the pieces resulting from a fragmentation event will be smaller than n_0. (For example, suppose $n_0 = 3$ and an aggregate of size 6 is going to be fragmented. There are five sites, each of which is equally likely to be chosen. If the middle fragmentation site is chosen, the resulting pieces will each be 3, the same as n_0. Every other fragmentation site will result in one piece that is smaller than n_0, which will return to the monomer state, and one piece larger than n_0, which will remain an aggregate. The two outermost fragmentation sites will result in recovering one monomer, and the last two will recover two monomers. The sum of monomers recovered, regardless of the original aggregate size, will always be $n_0(n_0 - 1)$.

As such, the total "rate in" portion of the differential equation for $x(t)$ is as follows,

$$\text{Rate In} = \alpha + \gamma n_0 (n_0 - 1) y_{n_0}(t) + \gamma n_0 (n_0 - 1) y_{n_0 + 1}(t) + \dots$$

$$:= \alpha + \gamma n_0 (n_0 - 1) \sum_{i \geq n_0} y_i(t). \tag{Eq. 6}$$

Combining these together, we have the differential equation for $x(t)$ as follows,

$$\frac{dx(t)}{dt} = \underbrace{\left(\alpha + \gamma n_0(n_0 - 1) \sum_{i \geq n_0} y_i(t) \right)}_{\text{Rate In}} - \underbrace{\left(\mu x(t) + 2\beta x(t) \sum_{i \geq n_0} y_i(t) \right)}_{\text{Rate Out}}.$$

$$\tag{Eq. 7}$$

Similar reasoning results in the following differential equations for $y_{n_0}(t)$ and $y_i(t)$ when $i > n_0$,

$$\frac{dy_{n_0}(t)}{dt} = \underbrace{\left(2\gamma \sum_{j > n_0} y_j(t) \right)}_{\text{Rate In: Bigger Aggregates Fragment}}$$

$$- \left(\underbrace{2\beta x(t) y_{n_0}(t)}_{\text{Conversion}} + \underbrace{\gamma(n_0 - 1)y_{n_0}(t)}_{\text{Fragmentation}} + \underbrace{\mu y_{n_0}(t)}_{\text{Degradation}} \right), \tag{Eq. 8}$$

$$\underbrace{\phantom{2\beta x(t) y_{n_0}(t) + \gamma(n_0 - 1)y_{n_0}(t) + \mu y_{n_0}(t)}}_{\text{Rate Out}}$$

$$\frac{dy_i(t)}{dt} = \left(\underbrace{2\beta x(t) y_{i-1}(t)}_{\text{Conversion}} + \underbrace{2\gamma \sum_{j > i} y_j(t)}_{\text{Fragmentation}} \right)$$

$$\underbrace{\phantom{2\beta x(t) y_{i-1}(t) + 2\gamma \sum_{j>i} y_j(t)}}_{\text{Rate In}}$$

$$- \left(\underbrace{2\beta x(t) y_i(t)}_{\text{Conversion}} + \underbrace{\gamma(i - 1)y_i(t)}_{\text{Fragmentation}} + \underbrace{\mu y_i(t)}_{\text{Degradation}} \right). \tag{Eq. 9}$$

$$\underbrace{}_{\text{Rate Out}}$$

Now that we have our mathematical model, we can begin the process of analyzing it. Generally, there are two approaches: analytical and numerical. With analytical approaches, we generally use the structures of the equations themselves to determine key characteristics, such as steady states (cases where the system configuration will remain unchanged) and their stability (where the system will tend over time). With numerical approaches, we use software (Matlab, R, Mathematica, Python, etc.) to study the temporal dynamics of the biochemical species in the system. To use numerical methods, we typically need to specify the initial condition of the system and all biochemical reaction rates (α, β, γ, μ, n_0). We will consider each approach below to answer the two questions we posed originally.

Stability of the prion aggregate system: Analytical approach

Our model (Equations 7–9) may appear quite daunting, but in fact is highly amenable to analytical methods. We will use these methods to consider our first question about what causes prion aggregates to be cleared and the disease phenotype to be reversed. We first rearrange the equations into a form that makes them easier to work with. Rather than considering each aggregate size y_i, we will consider two new quantities about the aggregates as follows:

$$Y(t) = \sum_{i \geq n_0} y_i(t), \tag{Eq. 10}$$

and

$$Z(t) = \sum_{i \geq n_0} i y_i(t). \tag{Eq. 11}$$

Note that $Y(t)$ represents the total concentration of prion aggregates, whereas $Z(t)$ represents the concentration of total protein in prion aggregates. Remarkably, our system of differential equations in y_i can be arranged into a set of differential equations for $Y(t)$, $Z(t)$, and $x(t)$ (our original monomer population),

$$\frac{dx(t)}{dt} = \alpha + \gamma n_0(n_0 - 1)Y(t) - \mu x(t) - 2\beta x(t)Y(t) \tag{Eq. 12}$$

$$\frac{dY(t)}{dt} = \gamma Z(t) - (\mu + \gamma(2n_0 - 1))Y(t) \tag{Eq. 13}$$

$$\frac{dZ(t)}{dt} = 2\beta x(t)Y(t) - \mu Z(t) - \gamma n_0(n_0 - 1)Y(t). \tag{Eq. 14}$$

With only three equations, it is much easier to solve for the steady states of the system. Steady states represent cases where the system is unchanging, so we identify them by finding values of x, Y, and Z that satisfy the following,

$$\frac{dx}{dt} = \frac{dY}{dt} = \frac{dZ}{dt} = 0. \tag{Eq. 15}$$

Some algebra will demonstrate that the system has exactly two steady-state solutions. The first is the aggregate-free steady state: $x = \alpha/\mu$, $Y = Z = 0$. In this case, all protein is in the healthy monomer state. The second nontrivial steady state (x^*, Y^*, Z^*) is more clearly shown in the following relationships,

$$x^* = \frac{\alpha/\mu}{R_0} \tag{Eq. 16}$$

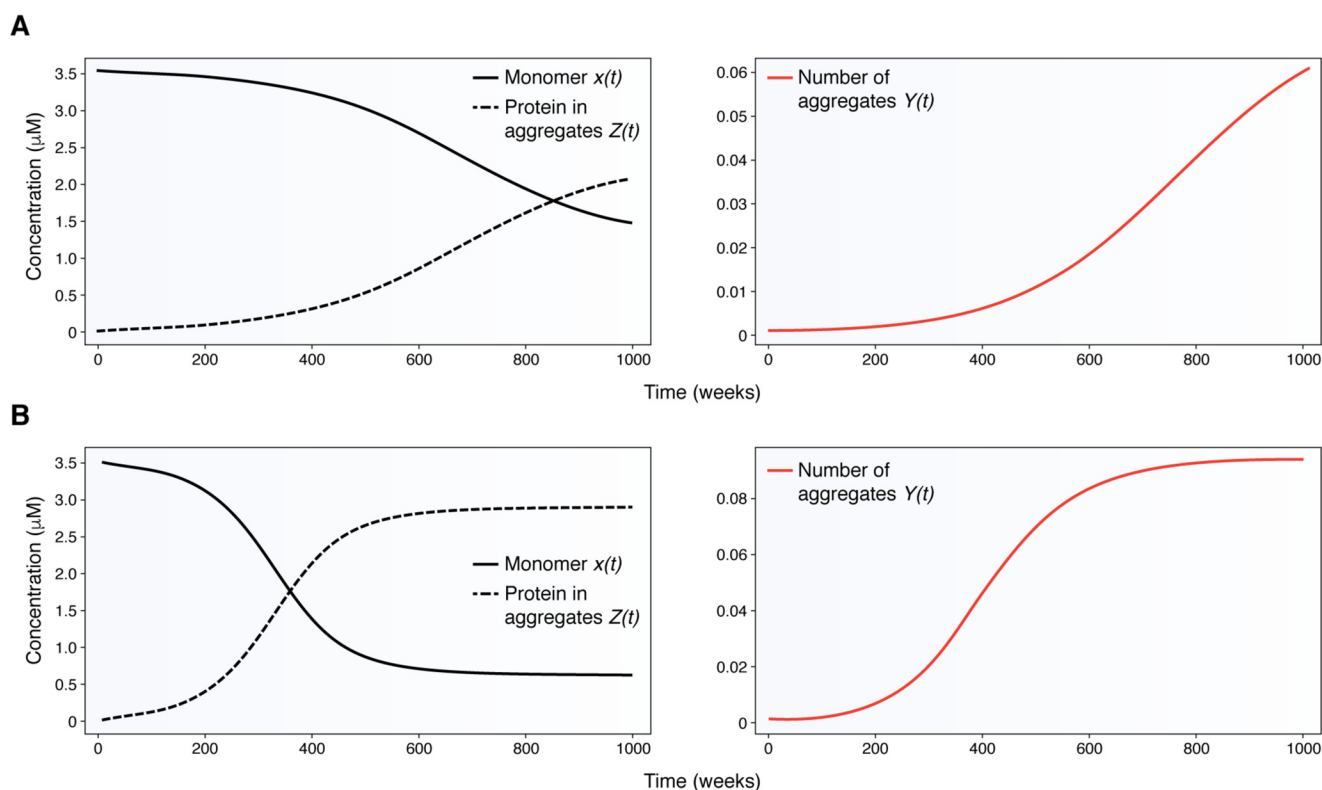$$Z^* + x^* = \alpha/\mu \tag{Eq. 17}$$

$$Z^*/Y^* = 2n_0 - 1 + \frac{\mu}{\gamma} \tag{Eq. 18}$$

where

$$R_0 = \frac{2\alpha\beta\gamma}{\mu(\mu + \gamma(n_0 - 1))(\mu + \gamma n_0)}. \tag{Eq. 19}$$

A

B

Figure 5. Temporal kinetics of prion aggregation. We plot the temporal evolution of the concentration of monomer $x(t)$ and protein in aggregates $Z(t)$ (*left*) and the number of aggregates $Y(t)$ (*right*) for "theoretical" choice of parameters. We note that although prion aggregates are present at the initial condition, there is a long delay before we see a significant decrease in the healthy protein. In *B*, the conversion rate is doubled, causing the time it takes for the system to reach steady state to be cut in half.

Those familiar with epidemiology will recognize the use of the R_0 value to represent the basic reproductive number (21). In epidemiology, R_0 represents the number of secondary infections produced by one primary infection in a susceptible population. When $R_0 < 1$, no matter how many aggregates are initially present, they will all eventually be cleared, and the system will converge to the aggregate-free steady state. When $R_0 > 1$, the introduction of any initial amount of aggregate will lead the system to converge to the nontrivial steady state.

How does this analytical work help us to better understand how biochemical rates can be manipulated to clear prion aggregates? Deriving an algebraic expression such as the one given in Equation 19 above is key. We know that when R_0 changes from a value greater than 1 to a value smaller than 1, prion aggregates will be completely cleared. But what parameter values should we change in order to arrive at such an R_0 value? And by how much? Whereas it might seem obvious that increasing the degradation rate μ above a certain threshold will cause prion aggregates to be cleared, lack of an algebraic expression for R_0 would make it almost impossible to determine the precise degree of change in parameters to produce the desired output.

For example, the parameters used to plot the system in Fig. 5A produce an R_0 value of $\sim$5.7. As expected, because this R_0 value is greater than 1, prion aggregates do persist and reach a positive stable steady state as shown. However, if we fix all other parameters and only change the fragmentation rate γ, we can use Equation 19 above to show how big or small γ must be to produce an R_0 value of <1 where prion aggregates are cleared.

Keeping all parameters the same as in Fig. 5A, we have the following,

$$R_0 < 1 \Rightarrow \frac{2\alpha\beta\gamma}{\mu(\mu + \gamma(n_0 - 1))(\mu + \gamma n_0)} < 1 \quad \text{(Eq. 20)}$$

$$\Rightarrow \gamma > 0.00330561.$$

Dynamics of the prion aggregate system: Numerical approach

To answer our second question, why there is such a long lag time in the manifestation of prion disease, we consider numerical simulations of our prion model. Whereas at one point, the lack of easy-to-use computational tools made it difficult for all but a relatively small group of experts to numerically solve differential equations, this is no longer the case. There are a multitude of free-to-use computational tools that enable anyone to numerically analyze even sophisticated mathematical models. Of course, as the complexity of the model grows, typically more effort is required to numerically code the solution. In this work, we use an iPython Notebook to analyze the dynamics of aggregates with an eye toward understanding why it takes so long for prion diseases to manifest.

In Fig. 5, we illustrate the concentration of $x(t)$ and $Z(t)$ (*A* and *B*, *left*) as well as the aggregate concentration $Y(t)$ (*A* and *B*, *right*) for a particular choice of biochemical parameters upon the introduction of a very small amount of prion aggregate $Z(0) = Y(0) = 1$ nM. (The biochemical parameters were chosen to roughly mirror the properties of the [PSI$^+$] weak strain in

yeast but plotted on a time scale relative to mammalian prion disease (40, 49).) In *A*, we see that the amount of healthy monomer $x(t)$ decreases relatively slowly; after 400 weeks, only 9% of the protein has changed to the prion (misfolded) state. This is because, initially, the aggregates are at a very low concentration, and thus conversion of monomer is favored over fragmentation. Because conversion only adds to existing aggregates (but does not create them), this has only a modest impact on the pool of healthy protein. However, just because we cannot see much of a decrease in $x(t)$ does not mean we are safe!

Eventually, the aggregates increase enough in size such that fragmentation becomes significant enough to create new aggregates at a discernible rate. This autocatalytic process then increases the rate of aggregate conversion and eventually produces a noticeable decrease in the healthy protein. It is precisely this effect, the initial balance that favors conversion over fragmentation, that causes the long-term delay in the manifestation of prion disease phenotypes. In *B*, we doubled the conversion rate, and as a result, the system reached its steady state almost twice as fast! We next describe how this model was used by us and previous researchers to learn about prion aggregation processes.

Interaction between mathematical models and experiments

With a mathematical model in place, it is now time to begin the process of using it in concert with experiments to study and probe the system. One common first step is to fit parameters by relating model output to experimentally observable quantities. For example, when the NPM was introduced, the authors related the biochemical parameters (α, β, γ, μ, n_0) to the exponential growth rate in aggregated protein they observed during early phases of prion disease through the relationship of the parameters to the R_0 (44). Similarly, the mathematicians Greer, Pujo-Menjouet, and Webb used prior experimental studies to fit the kinetic parameters of their model to a strain of scrapie (46, 50). With fit kinetic parameters, models can then make predictions about what should happen under particular initial conditions, $X(0)$, $Y(0)$, $Z(0)$. Indeed, as protein aggregation modeling has advanced, easy-to-use computational pipelines have been developed for researchers interested in fitting their *in vitro* aggregation curves (51).

When a model does not match experiments, this typically means the mathematical representation is not considering all of the relevant behavior of the biological system. Deciding what is missing and how to modify the model is challenging and requires the close interaction of experimental and mathematical scientists. However, this is when models can be their most useful, because they can suggest novel hypotheses and experiments. Researchers found that the NPM did not consistently match experimental data for the [PSI^+] prion (49). The only way the authors found to simultaneously fit all experimental observations (steady-state levels of soluble protein, average aggregate size, and loss of the prion phenotype) was to add two features to the model: 1) aggregate transmission during cell division is biased by aggregate size, and 2) fragmentation depends on a rate-limiting molecular chaperone. This suggested new experiments that would only be true if the new model was true. First, if aggregate transmission is size-dependent, the mathematical model predicted that the level of soluble protein a cell has should be linked to its age. This prediction was then experimentally validated, by sorting cells by soluble protein and counting their bud scars (generational age). Second, if fragmentation is rate-limiting, then increasing the synthesis rate α was predicted by the mathematical model to increase the size of aggregates. (Under the original NPM, increasing the synthesis rate α does not change the shape of the aggregate distribution.) This prediction was also experimentally validated by using a different promoter.

Discussion

We have now seen two examples about building mathematical models with our three-step process and highlighted how they can be used to learn about the molecular processes directing prion aggregation in yeast. In this section, we aim to provide more insight about how to introduce added complexity into a modeling framework. Specifically, we will discuss several examples of models of protein aggregation dynamics that have been extended to incorporate processes at distinct biological scales. In addition, the end of this section will discuss some of the challenges in modeling and give additional comments on considerations in using mathematical models.

One challenge in mathematical modeling is to determine the right type of model to answer the scientific questions being posed. Often researchers understand that a mathematical or computational model can be a valuable tool, but they may seek to develop a model before establishing exactly what questions they want to answer. The choice of model depends on many factors. Starting with several well-defined questions or hypotheses can help determine both the scales and level of complexity that need to be considered and lead to an appropriate choice of model. For example, some questions that might be asked include the following.

- What is the catalytic rate of a particular enzyme reaction? To address this question, a model must minimally include processes at the molecular scale. However, depending on desired complexity, the choice could be made to model the reaction of interest as a one-step process or include several successive steps in the reaction. In addition, important steps impacting an enzyme reaction could occur at the same scale, or some step might occur on a larger scale, such as cell behaviors or nutrient gradients that impact the environment where the reaction is taking place.

- Why does my experimental system have such high variance? Sensitivity analysis tools offer great methods to identify the cause of variance in complex biological systems. To answer this question, a model could consider one scale only and investigate the impact of variability of one or two factors at the same scale. Or the model could incorporate processes at the molecular scale and processes at the cell or tissue scale and use global sensitivity analysis methods to identify which factor(s) has the most impact on the overall variance of the system.

- What are the best time points to sample my experimental system to observe dynamics of the behavior I wish to study? Biological processes in the same system happen at
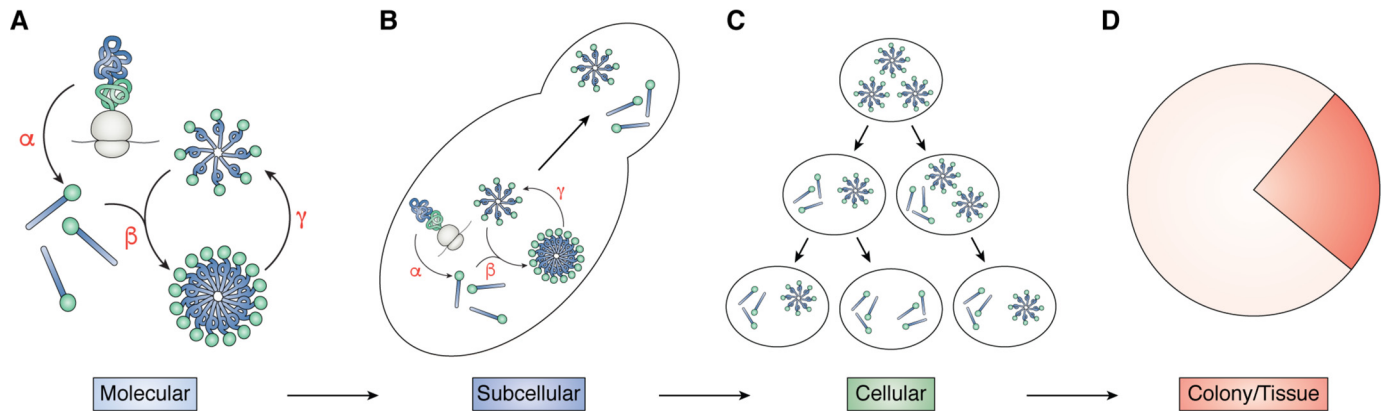
Figure 6. Multiscale nature of biological systems. A diagram depicts the multiscale nature of prion disease dynamics. *A*, biochemical scale. Shown are the key biochemical players and their interactions. Protein is synthesized at a rate α, monomers are converted into aggregates at a rate β, and fragmentation of aggregates occurs at a rate γ. *B*, subcellular scale. The process of conversion, fragmentation, synthesis, and degradation of protein aggregates occurs inside individual cells. The presence of molecular chaperones, protein degradation factors, and other substances varies inside each individual cell and can impact the rate of biochemical interactions. *C*, multicellular scale. Division in yeast occurs through budding, a process during which protein and aggregates are segregated between the mother and daughter cell. Modeling individual cell behaviors makes it possible to study the interplay of molecular, subcellular, and multicellular phenomena. *D*, colony/tissue scale. At the largest scale, prion disease in yeast manifests as different phenotypes at the colony level (*white*, disease; *red*, disease-free), the most interesting of which is sectored colonies where sections of cells within a majority diseased colony lose all of their aggregates and become disease-free (84). Moreover, multiscale models hope to connect important components at the biochemical scale with observed spatiotemporal patterns in colony phenotypes. In mammalian neurodegenerative diseases, disease phenotypes are observed at the level of an organ (85–86) (2).

different time scales. For example, chemical reactions happen much faster than the time it takes for a cell to divide or move across an agar plate. Depending on the scientific question being asked, samples at different time points can provide necessary experimental data to calibrate model components at different scales.

When scientific questions lead model development, it becomes more clear what processes and scales must be included in the mathematical modeling framework, and this determines the kind of mathematical model and analysis possible.

Incorporating processes at multiple biological scales

Biological systems are inherently multiscale, because all living organisms are composed of many different functional networks that operate across diverse temporal and spatial scales (Fig. 6). Technological advances have made it possible to attain highly detailed descriptions of key biological components at almost any scale imaginable (*i.e.* genetic mutations, formation of a blood clot, and even the growing and shortening phases of microtubules in individual cells) (52–57). Because cutting edge technology makes it possible to observe biological phenomena at each unique layer of organization, our understanding of how life works is no longer limited by the instruments in our laboratories. However, with every new experiment comes additional complexity. The abundance of information available has ultimately resulted in the compartmentalization of scientific inquiry and led to separate study of each different biological scale (*i.e.* molecular biologists *versus* cellular, organismic, or population, etc.). Mathematical and computational models provide a tool to integrate knowledge from different scales and identify how collective interactions on a fundamental scale can give rise to large-scale phenomena (Fig. 6). Thus, one challenge in building a mathematical model is determining on what scale (or scales) the key players and interactions occur. One important question to ask is which scale(s) of resolution will provide

the most information for understanding the underlying mechanisms of the system? An equally important and converse question is whether there is a scale of resolution that offers no insight and should be simplified or ignored.

To further illustrate how to incorporate biological processes at different scales into a modeling framework, we consider the multiscale nature prion disease dynamics (Fig. 6). At the most basic scale, biochemical interactions depicting conversion, fragmentation, synthesis, and degradation form the basic mechanisms that lead to the presence of prion disease (Fig. 6A). For this reason, the majority of mathematical models developed to answer the question of what causes prion aggregates to be cleared have only considered interactions between components on the molecular scale (see Section 3). However, in some instances, these models were not able to reproduce experimental data. In this case, the given modeling framework can be extended to include processes at the next level of organization in order to attempt to produce model results that agree with experimental data (Fig. 6B).

For yeast prions, the next level of organization is subcellular. Because molecular chaperones have been shown to be essential for prion propagation (49, 58), Davis *et al.* (59) extended the NPM by modifying the existing biochemical interactions given in Section 3 to account for the impact of molecular chaperones. The addition of these new biochemical interactions resulted in better agreement with experimental results that could not be supported by the original equations alone. In addition, this work further supported the hypothesis that there exist many different prion strains. In many cases, the distinction between models that look only at the molecular and/or biochemical scale and models that incorporate processes on the subcellular scale involves adding differential equations to an already existing modeling framework to describe processes such as diffusion or spatially varying density functions in addition to keeping track of individual particles and/or groups of particles.

The next tier of resolution is the whole cell (Fig. 6C). One of the most unique complications introduced by studying prion disease in yeast comes with the consideration that prions in yeast propagate within a colony of growing and dividing cells. The time it takes for protein to change configurations is much faster than the time it takes for a cell to grow or divide. Thus, an interesting question to consider is how cellular behaviors such as cell cycle length, protein segregation at the time of division, and/or age of cells impact protein aggregation dynamics throughout the entire colony (Fig. 6, *B* and *C*). To address this question, a new model must include interactions between intracellular components within each individual cell, individual cell behaviors impacting intracellular dynamics, and cell-cell interaction between many different cells in the same colony. One type of well-established modeling framework that is capable of integrating molecular, subcellular, and cellular level processes is agent-based models. Agent-based models represent cells as discrete units that interact with each other and can also carry out individual cell processes such as protein aggregation, division, and growth (for reviews, see Refs. 60–69) Fortunately, incorporating processes at different scales using agent-based models does not require starting from scratch! In the case of yeast prion biology, one of the many well-studied models of protein aggregation that focus on the molecular scale only can be coupled with an agent-based framework. Thus, developing a model that incorporated processes at the cellular scale would only require building an agent-based framework or relying on one of the many great code bases that exist in the literature.

At the highest scale, prion disease manifests as tissue atrophy in mammals and population level phenotypes in yeast (Fig. 6D). In many studies, population level data are the only quantitative output available from experiments. However, the concentration of a protein or other cellular constituent is known to vary considerably among cells in the same colony. This heterogeneity is thought to arise from several sources, including differences in kinetic rates between individual cells and distribution of cellular constituents at the time of cell division. Connecting interactions between components at the fundamental scale with population level phenotypes is experimentally challenging. However, in multiscale models, perturbations of parameters at the fundamental scale (*i.e.* protein modifications) can generate observable and measurable changes to coarse-grained outputs at the population level (*i.e.* colony phenotype). Integrating processes across different spatial and temporal scales can be achieved using agent-based models described earlier. Furthermore, agent-based models have been used in many applications, including tumor growth, blood clot formation, stem cell regulation in plants and understanding the interplay of biochemical and molecular parameters on individual cell behaviors. Another class of model that can be used to study dynamics of entire cell colonies, tissues, and organs are continuous models that use ordinary differential equations or partial differential equations to represent the growth of a tissue as one continuous sheet or the change in the shape or position over time of the edge of a colony or group of cells as one continuous boundary. In many cases, discrete, agent-based modeling frameworks have been used to derive differential equation

models that can approximate large-scale behavior more efficiently or infer parameters for large-scale behavior models.

Inherent to the presence of different spatial and time scales within biological systems, another significant distinction occurs between data from *in vitro* and *in vivo* experiments. A large amount of biological data, particularly at the molecular level, is obtained from *in vitro* experiments. Due to experimental complexity, observations are often restricted to single spatial and/or temporal scales. As a result, observations from *in vitro* and *in vivo* experiments can have very different outcomes. In the case of prion disease, aggregates within an organism are fragmented by chaperones, protein degradation factors are present, and protein is continually being synthesized. However, in *in vitro* assays, fragmentation necessarily operates without the complete cellular machinery and is typically operating under very different concentrations. Thus, building a model that can resolve the discrepancy between *in vitro* and *in vivo* experiments by encompassing *in vivo* processes that may be missing at the *in vitro* scale can prove very helpful. Mathematical and computational models are uniquely positioned to capture the connectivity between these divergent scales of biological function and have the potential to bridge the gap between isolated *in vitro* experiments at the most basic scale and whole organism *in vivo* models with organism or population level output.

Choosing a mathematical framework

There are many different mathematical tools that can be used to convert the interactions at each scale given by the conceptual model into a mathematical framework (20, 70). One main distinction in mathematical frameworks is deterministic *versus* stochastic. In Section 3, we considered a deterministic model of prion aggregation. In other words, given an identical set of initial conditions, the model will always produce the same output. This is in contrast to a stochastic model, in which we consider the evolution of the probability that the system occupies any particular state. Intriguingly, the same set of biochemical equations (see Fig. 4 (*middle*)) can be translated into either a stochastic or deterministic framework with the law of mass action (*e.g.* see Refs. 17 and 18). Ultimately, the scientific question will determine the precise mathematical framework to be used. For example, if we are interested in the probability of the appearance of a prion phenotype, we need to consider a stochastic model, as has been done by many (*e.g.* see Refs. 71 and 72).

In addition to deterministic *versus* stochastic frameworks, there are many different options for the types of equations that can be used inside each type of model. For example, as detailed in Section 3, systems of differential equations using the law of mass action kinetics can be leveraged to represent chemical reactions. In addition, systems of differential equations (ordinary or partial) are also ideal for describing the concentrations of signaling molecules in both intracellular (inside one cell) and extracellular (moving throughout many cells) domains.

Another choice to be made is whether to build a continuous or discrete mathematical framework. In many applications, one multiscale phenomenon we wish to model is how the interaction between intracellular processes, such as protein aggrega-

tion, happening inside each individual cell within a tissue or colony, and the individual cell behaviors lead to emergent patterns or phenotypes at the organismal or population level. In such cases, it is helpful to treat the intracellular process happening inside each cell as a continuous variable using differential equations and model individual cells as discrete entities that interact. Models that represent cells as discrete entities are generally referred to as agent-based or cell-based models, and this class of model has been used in many different applications (for reviews, see Refs. 60 –67 and 73–78).

Conclusions

There has never been a better time for interdisciplinary research collaborations between experimental and mathematical scientists. Although communicating across domains is still challenging, the potential benefit from using mathematical models in new settings is significant. Moreover, with the increasing availability of user-friendly software, it is easier for scientists of all backgrounds to develop and analyze their own mathematical models (79 –83). It is our hope that the three-step pipeline we have identified will enable readers to pursue developing their own mathematical models and spur productive discussions with mathematical scientists. Finally, we recognize that developing a new mathematical model is always challenging, and for brevity, our review contains only one modeling scenario. We encourage interested readers to refer to the many textbooks that we have found valuable in our own learning and teaching (16 –22) or to one of the other reviews in building mathematical models (12–15).

References

1. Cohen, J. E. (2004) Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PLoS Biol.* **2,** e439 CrossRef Medline
2. Thompson, P. M., Hayashi, K. M., de Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., Herman, D., Hong, M. S., Dittmer, S. S., Doddrell, D. M., Toga, A. W. (2003) Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* **23,** 994 –1005 CrossRef Medline
3. Hunter, P. (2010) Biology is the new physics. *EMBO Rep.* **11,** 350–352 CrossRef Medline
4. McKenna, M. T., Weis, J. A., Brock, A., Quaranta, V., and Yankeelov, T. E. (2018) Precision medicine with imprecise therapy: computational modeling for chemotherapy in breast cancer. *Transl. Oncol.* **11,** 732–742 CrossRef Medline
5. Tang, L., van de Ven, A. L., Guo, D., Andasari, V., Cristini, V., Li, K. C., and Zhou, X. (2014) Computational modeling of 3D tumor growth and angiogenesis for chemotherapy evaluation. *PLoS One* **9,** e83962 CrossRef Medline
6. Garraway, L. A., Verweij, J., and Ballman, K. V. (2013) Precision oncology: an overview. *J. Clin. Oncol.* **31,** 1803–1805 CrossRef Medline
7. Olopade, O. I., Grushko, T. A., Nanda, R., and Huo, D. (2008) Advances in breast cancer: pathways to personalized medicine. *Clin. Cancer Res.* **14,** 7988 –7999 CrossRef Medline
8. Altrock, P. M., Liu, L. L., and Michor, F. (2015) The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer* **15,** 730–745 CrossRef Medline
9. Yankeelov, T. E., Atuegwu, N., Hormuth, D., Weis, J. A., Barnes, S. L., Miga, M. I., Rericha, E. C., and Quaranta, V. (2013) Clinically relevant modeling of tumor growth and treatment response. *Sci. Transl. Med.* **5,** 187ps9 CrossRef Medline
10. Anderson, A. R. A., and Quaranta, V. (2008) Integrative mathematical oncology. *Nat. Rev. Cancer* **8,** 227–234 CrossRef Medline
11. Yankeelov, T. E., An, G., Saut, O., Luebeck, E. G., Popel, A. S., Ribba, B., Vicini, P., Zhou, X., Weis, J. A., Ye, K., and Genin, G. M. (2016) Multi-scale modeling in clinical oncology: opportunities and barriers to success. *Ann. Biomed. Eng.* **44,** 2626 –2641 CrossRef Medline
12. Torres, N. V., and Santos, G. (2015) The (mathematical) modeling process in biosciences. *Front. Genet.* **6,** 354 CrossRef Medline
13. Tomlin, C. J., and Axelrod, J. D. (2007) Biology by numbers: mathematical modelling in developmental biology. *Nat. Rev. Genet.* **8,** 331–340 CrossRef Medline
14. Ay, A., and Arnosti, D. N. (2011) Mathematical modeling of gene expression: a guide for the perplexed biologist. *Crit. Rev. Biochem. Mol. Biol.* **46,** 137–151 CrossRef Medline
15. Chen, W. W., Niepel, M., and Sorger, P. K. (2010) Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* **24,** 1861–1875 CrossRef Medline
16. Alon, U. (2006) *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall/CRC, London
17. Allen, L. J. S. (2010) *An Introduction to Stochastic Processes with Applications to Biology*, Chapman and Hall/CRC, London
18. Wilkinson, D. J. (2011) *Stochastic Modelling for Systems Biology*, Chapman and Hall/CRC, London
19. Allen, L. J. S. (2007) *Introduction to Mathematical Biology*, Pearson/Prentice Hall, London
20. Edelstein-Keshet, L. (2005) *Mathematical Models in Biology*, SIAM, Philadelphia
21. Brauer, F., Castillo-Chavez, C., and Castillo-Chavez, C. (2001) *Mathematical Models in Population Biology and Epidemiology*, Vol. 40, Springer, New York
22. Segel, L. A., and Edelstein-Keshet, L. (2013) *A Primer in Mathematical Models in Biology*, Vol. 129, SIAM, Philadelphia
23. Michaelis, L., and Menten, M. L. (1913) The kinetics of the inversion effect. *Biochem. Z.* **49,** 333–369
24. Brown, P., Will, R. G., Bradley, R., Asher, D. M., and Detwiler, L. (2001) Bovine spongiform encephalopathy and variant Creutzfeldt-Jakob disease: background, evolution, and current concerns. *Emerg. Infect. Dis.* **7,** 6 –16 CrossRef Medline
25. Brettschneider, J., Del Tredici, K., Lee, V. M., and Trojanowski, J. Q. (2015) Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nat. Rev. Neurosci.* **16,** 109 –120 CrossRef Medline
26. Medori, R., Tritschler, H. J., LeBlanc, A., Villare, F., Manetto, V., Chen, H. Y., Xue, R., Leal, S., Montagna, P., and Cortelli, P. (1992) Fatal familial insomnia, a prion disease with a mutation at codon 178 of the prion protein gene. *N. Engl. J. Med.* **326,** 444 – 449 CrossRef Medline
27. Brotherston, J. G., Renwick, C. C., Stamp, J. T., Zlotnik, I., and Pattison, I. H. (1968) Spread of scrapie by contact to goats and sheep. *J. Comp. Pathol.* **78,** 9 –17 CrossRef Medline
28. Vilette, D., Courte, J., Peyrin, J. M., Coudert, L., Schaeffer, L., Andréoletti, O., and Leblanc, P. (2018) Cellular mechanisms responsible for cell-to-cell spreading of prions. *Cell. Mol. Life Sci.* **75,** 2557–2574 CrossRef Medline
29. Collinge, J., Whitfield, J., McKintosh, E., Beck, J., Mead, S., Thomas, D. J., and Alpers, M. P. (2006) Kuru in the 21st century—an acquired human prion disease with very long incubation periods. *Lancet* **367,** 2068 –2074 CrossRef Medline
30. Alper, T., Cramp, W. A., Haig, D. A., and Clarke, M. C. (1967) Does the agent of scrapie replicate without nucleic acid? *Nature* **214,** 764 –766 CrossRef Medline
31. Alper, T., Haig, D. A., and Clarke, M. C. (1966) The exceptionally small size of the scrapie agent. *Biochem. Biophys. Res. Commun.* **22,** 278 –284 CrossRef Medline
32. Griffith, J. S. (1967) Nature of the scrapie agent: self-replication and scrapie. *Nature* **215,** 1043–1044 CrossRef Medline
33. Pattison, I. H., and Jones, K. M. (1967) The possible nature of the transmissible agent of scrapie. *Vet. Rec.* **80,** 2–9 CrossRef Medline
34. Tuite, M. F., and Serio, T. R. (2010) The prion hypothesis: from biological anomaly to basic regulatory mechanism. *Nat. Rev. Mol. Cell Biol.* **11,** 823– 833 CrossRef Medline
35. Aguzzi, A., and Polymenidou, M. (2004) Mammalian prion biology: one century of evolving concepts. *Cell* **116,** 313–327 CrossRef Medline

ＳＡＳBMB

J. Biol. Chem. (2020) 295(15) 5022–5035 **5033**

36. Rhodes, R. (1999) *Deadly Feasts: Tracking the Secrets of a Terrifying New Plague*, Simon and Schuster, New York

37. Sindi, S. S., and Serio, T. R. (2009) Prion dynamics and the quest for the genetic determinant in protein-only inheritance. *Curr. Opin. Microbiol.* **12,** 623–630 CrossRef Medline

38. Sindi, S. S. (2017) Mathematical modeling of prion disease. In *Prion: An Overview*, pp. 208–227, IntechOpen, London

39. Ridout, M., Giagos, V., Morgan, B., Naeimi, W., von der Haar, T., and Tuite, M. (2011) Modelling prion dynamics in yeast. *Proceedings of the 58th World Statistical Congress*, Dublin (Session IPS020), International Statistical Institute, Voorberg, Netherlands

40. Tanaka, M., Collins, S. R., Toyama, B. H., and Weissman, J. S. (2006) The physical basis of how prion conformations determine strain phenotypes. *Nature* **442,** 585–589 CrossRef Medline

41. Carbonell, F., Iturria-Medina, Y., and Evans, A. C. (2018) Mathematical modeling of protein misfolding mechanisms in neurological diseases: a historical overview. *Front. Neurol.* **9,** 37 CrossRef Medline

42. Wang, H., Rhoads, D. D., and Appleby, B. S. (2019) Human prion diseases. *Curr. Opin. Infect. Dis.* **32,** 272–276 CrossRef Medline

43. Jaunmuktane, Z., Mead, S., Ellis, M., Wadsworth, J. D., Nicoll, A. J., Kenny, J., Launchbury, F., Linehan, J., Richard-Loendt, A., Walker, A. S., Rudge, P., Collinge, J., and Brandner, S. (2015) Evidence for human transmission of amyloid-β pathology and cerebral amyloid angiopathy. *Nature* **525,** 247–250 CrossRef Medline

44. Masel, J., Jansen, V. A., and Nowak, M. A. (1999) Quantifying the kinetic parameters of prion replication. *Biophys. Chem.* **77,** 139–152 CrossRef Medline

45. Xue, W.-F. (2015) Nucleation: the birth of a new protein phase. *Biophys. J.* **109,** 1999–2000 CrossRef Medline

46. Greer, M. L., Pujo-Menjouet, L., and Webb, G. F. (2006) A mathematical analysis of the dynamics of prion proliferation. *J. Theor. Biol.* **242,** 598–606 CrossRef Medline

47. Knowles, T. P. J., Vendruscolo, M., and Dobson, C. M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **15,** 384–396 CrossRef Medline

48. Xue, W.-F., and Radford, S. E. (2013) An imaging and systems modeling approach to fibril breakage enables prediction of amyloid behavior. *Biophys. J.* **105,** 2811–2819 CrossRef Medline

49. Derdowski, A., Sindi, S. S., Klaips, C. L., DiSalvo, S., and Serio, T. R. (2010) A size threshold limits prion transmission and establishes phenotypic diversity. *Science* **330,** 680–683 CrossRef Medline

50. Rubenstein, R., Merz, P. A., Kascsak, R. J., Scalici, C. L., Papini, M. C., Carp, R. I., and Kimberlin, R. H. (1991) Scrapie-infected spleens: analysis of infectivity, scrapie-associated fibrils, and protease-resistant proteins. *J. Infect. Dis.* **164,** 29–35 CrossRef Medline

51. Meisl, G., Kirkegaard, J. B., Arosio, P., Michaels, T. C., Vendruscolo, M., Dobson, C. M., Linse, S., and Knowles, T. P. (2016) Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Protoc.* **11,** 252–272 CrossRef Medline

52. Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., Hussain, M., Phillips, A. D., and Cooper, D. N. (2017) The human gene mutation database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum. Genet.* **136,** 665–677 CrossRef Medline

53. Gou, L., Bloom, J. S., and Kruglyak, L. (2019) The genetic basis of mutation rate variation in yeast. *Genetics* **211,** 731–740 CrossRef Medline

54. Alber, M., Xu, S., Xu, Z., Kim, O., Britton, S., Litvinov, R., and Weisel, J. (2019) Multi-scale models of deformation of blood clots. *Biophys. J.* **116,** 323a CrossRef

55. Kerssemakers, J. W., Munteanu, E. L., Laan, L., Noetzel, T. L., Janson, M. E., and Dogterom, M. (2006) Assembly dynamics of microtubules at molecular resolution. *Nature* **442,** 709–712 CrossRef Medline

56. Elliott, A., and Shaw, S. L. (2018) Update: plant cortical microtubule arrays. *Plant Physiol.* **176,** 94–105 CrossRef Medline

57. Hemmat, M., Castle, B. T., and Odde, D. J. (2019) Multi-scale computational modeling of tubulin-tubulin interactions in microtubule self-assembly from atoms to cells. *Biophys. J.* **116,** 256a CrossRef

58. Satpute-Krishnan, P., Langseth, S. X., and Serio, T. R. (2007) Hsp104-dependent remodeling of prion complexes mediates protein-only inheritance. *PLoS Biol.* **5,** e24 CrossRef Medline

59. Davis, J. K., and Sindi, S. S. (2016) A mathematical model of the dynamics of prion aggregates with chaperone-mediated fragmentation. *J. Math. Biol.* **72,** 1555–1578 CrossRef Medline

60. Anderson, A., Chaplain, M. A. J., and Rejniak, K. (eds) (2007) *Single-cell-based Models in Biology and Medicine*, Springer Science and Business Media, Basel, Switzerland

61. Fletcher, A. G., Cooper, F., and Baker, R. E. (2017) Mechanocellular models of epithelial morphogenesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372,** 20150519 CrossRef Medline

62. Sandersius, S. A., Weijer, C. J., and Newman, T. J. (2011) Emergent cell and tissue dynamics from subcellular modeling of active biomechanical processes. *Phys. Biol.* **8,** 045007 CrossRef Medline

63. Chaturvedi, R., Huang, C., Kazmierczak, B., Schneider, T., Izaguirre, J. A., Glimm, T., Hentschel, H. G., Glazier, J. A., Newman, S. A., and Alber, M. S. (2005) On multiscale approaches to three-dimensional modelling of morphogenesis. *J. R. Soc. Interface* **2,** 237–253 CrossRef Medline

64. Hwang, M., Garbey, M., Berceli, S. A., and Tran-Son-Tay, R. (2009) Rule-based simulation of multi-cellular biological systems—a review of modeling techniques. *Cell. Mol. Bioeng.* **2,** 285–294 CrossRef Medline

65. Paul Van Liedekerke, MM Palm, N Jagiella, and Dirk Drasdo. (2015) Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Comput. Part. Mech.* **2,** 401–444 CrossRef

66. Dirk Drasdo, Stefan Hoehme, and Michael Block. (2007) On the role of physics in the growth and pattern formation of multi-cellular systems: what can we learn from individual-cell based models? *J. Stat. Phys.* **128,** 287 CrossRef

67. Galle, J., Aust, G., Schaller, G., Beyer, T., and Drasdo, D. (2006) Individual cell-based models of the spatial-temporal organization of multicellular systems—achievements and limitations. *Cytometry A* **69,** 704–710 CrossRef Medline

68. Pathmanathan, P., Cooper, J., Fletcher, A., Mirams, G., Murray, P., Osborne, J., Pitt-Francis, J., Walter, A., and Chapman, S. J. (2009) A computational study of discrete mechanical tissue models. *Phys. Biol.* **6,** 036001 CrossRef Medline

69. Diaz de la Loza, M. C., and Thompson, B. J. (2017) Forces shaping the *Drosophila* wing. *Mech. Dev.* **144,** 23–32 CrossRef Medline

70. Murray, J. D. (2003) *Mathematical Biology I: An Introduction*, Vol. I, Springer, New York

71. Yvinec, R., D'Orsogna, M. R., and Chou, T. (2012) First passage times in homogeneous nucleation and self-assembly. *J. Chem. Phys.* **137,** 244107 CrossRef Medline

72. Davis, J. K., and Sindi, S. S. (2016) Initial condition of stochastic self-assembly. *Phys. Rev. E* **93,** 022109 CrossRef Medline

73. Deutsch, A., and Dormann, S. (2018) *Cellular Automaton Modeling of Biological Pattern Formation: Characterization, Examples, and Analysis*, Birkhäuser, New York

74. Deutsch, A., and Dormann, S. (2005) *Cellular Automaton Modeling of Biological Pattern Formation*, Springer, New York

75. Alber, M. S., Kiskowski, M. A., Glazier, J. A., and Jiang, Y. (2003) On cellular automaton approaches to modeling biological cells. In *Mathematical Systems Theory in Biology, Communications, Computation, and Finance* (Rosenthal, J., and Gilliam, D. S., eds) pp. 1–39, Springer, New York

76. Boon, J. P., Dab, D., Kapral, R., and Lawniczak, A. (1996) Lattice gas automata for reactive systems. *Phys. Rep.* **273,** 55–147 CrossRef

77. Chopard, B. (2012) *Cellular Automata Modeling of Physical Systems*, Springer, New York

78. Wolf-Gladrow, D. A. (2004) *Lattice-gas Cellular Automata and Lattice Boltzmann Models: An Introduction*, Springer, New York

79. Ghaffarizadeh, A., Heiland, R., Friedman, S. H., Mumenthaler, S. M., and Macklin, P. (2018) Physicell: An Open Source Physics-based Cell Simulator for 3-D Multicellular Systems. *PLoS Comput. Biol.* **14,** e1005991 CrossRef Medline

80. Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.-Y., and Kitano, H. (2011) Software for systems biology: from tools to integrated platforms. *Nat. Rev. Genet.* **12,** 821–832 CrossRef Medline

81. Herajy, M., Liu, F., Rohr, C., and Heiner, M. (2017) Snoopy's hybrid simulator: a tool to construct and simulate hybrid biological models. *BMC Syst. Biol.* **11,** 71 CrossRef Medline

82. Oliphant, T. E. (2006) *A Guide to NumPy*, 1st Ed., Trelgol Publishing

83. Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C. (2016) Jupyter notebooks—a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, (Loizides, F., and Schmidt, B., eds) pp. 87–90, IOS Press, Amsterdam, Netherlands

84. Klaips, C. L., Hochstrasser, M. L., Langlois, C. R., and Serio, T. R. (2014) Spatial quality control bypasses cell-based limitations on proteostasis to promote prion curing. *eLife* **3,** e04288 CrossRef Medline

85. Raj, A., Kuceyeski, A., and Weiner, M. (2012) A network diffusion model of disease progression in dementia. *Neuron* **73,** 1204–1215 CrossRef Medline

86. Apostolova, L. G., Steiner, C. A., Akopyan, G. G., Dutton, R. A., Hayashi, K. M., Toga, A. W., Cummings, J. L., and Thompson, P. M. (2007) Three-dimensional gray matter atrophy mapping in mild cognitive impairment and mild Alzheimer disease. *Arch. Neurol.* **64,** 1489–1495 CrossRef Medline

ASBMB

J. Biol. Chem. (2020) 295(15) 5022–5035 **5035**