bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The Roles of Space and Stochasticity in Computational Simulations of Cellular Biochemistry: Quantitative Analysis and Qualitative Insights

M. E. Johnson[1]*, A. Chen[1], J. R. Faeder[2], P. Henning[3], I. I. Moraru[4], M. Meier-Schellersheim[5], R. F. Murphy[6], T. Prüstel[5], J. A. Theriot[7], A. M. Uhrmacher[3]

[1]TC Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA

[2]Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA

[3]Institute of Computer Science, University of Rostock, Rostock, Germany

[4]Department of Cell Biology, Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT, USA

[5]Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

[6]Computational Biology Department, Department of Biological Sciences, Department of Biomedical Engineering, Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

[7]Department of Biology and Howard Hughes Medical Institute, University of Washington, Seattle, WA USA

*Corresponding author (all other authors are listed alphabetically)

Dr. Margaret E. Johnson
Johns Hopkins University
School of Arts & Sciences
121C Mergenthaler
3400 N. Charles Street
Baltimore, MD 21218

phone: 410-516-2376
email: margaret.johnson@jhu.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ABSTRACT

Most of the fascinating phenomena studied in cell biology emerge from interactions among highly organized multi-molecular structures and rapidly propagating molecular signals embedded into complex and frequently dynamic cellular morphologies. For the exploration of such systems, computational simulation has proved to be an invaluable tool, and many researchers in this field have developed sophisticated computational models for application to specific cell biological questions. However it is often difficult to reconcile conflicting computational results that use different simulation approaches (for example partial differential equations versus particle-based stochastic methods) to describe the same phenomenon. Moreover, the details of the computational implementation of any particular algorithm may give rise to quantitatively or even qualitatively different results for the same set of starting assumptions and parameters. In an effort to address this issue systematically, we have defined a series of computational test cases ranging from very simple (bimolecular binding in solution) to moderately complex (spatial and temporal oscillations generated by proteins binding to membranes) that represent building blocks for comprehensive three-dimensional models of cellular function. Having used two or more distinct computational approaches to solve each of these test cases with consistent parameter sets, we generally find modest but measurable differences in the solutions of the same problem, and a few cases where significant deviations arise. We discuss the strengths and limitations of commonly used computational approaches for exploring cell biological questions and provide a framework for decision-making by researchers wishing to develop new models for cell biology. As computational power and speed continue to increase at a remarkable rate, the dream of a fully comprehensive computational model of a living cell may be drawing closer to reality, but our analysis demonstrates that it will be crucial to evaluate the accuracy of such models critically and systematically.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

INTRODUCTION

21st-century cell biology has been transformed by rapid development of new technologies that have delivered our field into an era where scientists can now easily generate vast amounts of quantitative data, providing a broad and comprehensive view of problems that were previously accessible only by brief, laboriously achieved glances through tiny chinks. Northern blots have been replaced by RNA sequencing, Western blots are being replaced by proteomics, and the central cell biological tool of imaging has been revolutionized by a wealth of dramatic improvements in labeling methods, optical design, and digital imaging. While currently available data are just a tiny fraction of the amount that will eventually be needed to model entire cells, sufficient data is available for a number of cell biological processes to ask whether and how we can use it to understand the biological mechanisms underlying those processes.

Mathematical modeling can help (Cohen, 2004; Gunawardena, 2014). By generating a model of a process that we want to study, where we are able to define and control all the inputs and parameters, we can directly determine whether a mechanism we have hypothesized is actually sufficient to explain the phenomenon we have observed. While mathematical modeling of cell-biological processes can take many forms, here we will focus exclusively on computational simulation. One of our goals is to discuss the current state-of-the-art in computational cell biology and remaining open challenges, particularly regarding the spatial organization of signaling processes, the inclusion of stochastic effects, and the multi-scale nature of many cell-biological processes.

Computers are much better than humans at keeping track of large, complex systems with many interacting parts, and also at mercilessly following predetermined rules. One sub-field of cell biology that has been able to make great use of these abilities is the study of signal transduction, where typically many different molecular species interact with one another in branching and reticular networks. Biochemical and genetic experiments have been able to map out and characterize many individual pairwise interactions in multicomponent signaling pathways. However, the resulting “wiring diagrams” have provided little direct insight into how these networks of molecular interactions could give rise to the diverse and fascinating outputs of these systems, which might be able to generate oscillations, high-pass or low-pass filtering of receptor-mediated input signals, conversion of analog signals into switch-like binary outputs, etc. Computational simulations have been instrumental in bringing order and insight into this tangled web so that now it is possible to recognize recurrent motifs in the design of signal transduction systems, and, sometimes, accurately predict cellular responses to external stimuli (Cao et al., 2016; Eungdamrong and Iyengar, 2004; Janes and Lauffenburger, 2013; Kestler et al., 2008).

Within this context, a very fertile ecosystem of computational tools for modeling cellular biochemistry has flourished (Bartocci and Lio, 2016; Gillespie, 1977a).There are now also approaches that enable scientists who are not computational experts themselves to translate their hypotheses about cellular signaling mechanisms into formal models in compact and intuitive ways; these include using textual rules (Faeder, 2011; Harris et al., 2016; Maus et al., 2011) or iconographic symbols (Schaff et al., 2016; Sekar et al., 2017; Zhang et al., 2013) to specify molecular interactions. Based on those specifications, the tools generate the resulting computational representations of the signaling networks and allow modelers to easily modify their assumptions to explore the consequences of such simulated manipulations on cellular behavior. To allow for tool-independent formulation and sharing of computational models, the widely-used “systems biology markup

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

language" SBML was developed. SBML is continuously evolving and is supported by a large number of software systems for simulation and data analysis (Gillespie, 1976a; Gillespie, 1976b; Gillespie, 1977a; Gillespie, 1977b; SBML.org, 2020). While there are many variations, generally simulations of this kind are able to keep track of concentrations and interactions of many individual molecular species as they change over time, often the most interesting dimension for the study of signal transduction.

However, many cell biological processes, including some kinds of signaling, cannot be analyzed without the notion of space. Cell-cell communication is frequently based on the exchange of soluble messenger molecules, such as hormones, cytokines or chemokines, that diffuse through extracellular space before being captured by specific receptors at particular locations on cellular membranes. Direct cell-cell contacts also typically involve only a few of the receptors on a cell and generate localized signals that activate cascades of protein interactions and modifications to propagate from the membrane into the cytoplasm. Spatial simulation of cell biological phenomena is not new; in 1952, Hodgkin and Huxley simulated the propagation of an action potential down a neuronal axon (Gillespie, 1977a; Hodgkin and Huxley, 1952), and in the same year Alan Turing used computational simulation to demonstrate how chemical systems featuring both diffusion and reaction could generate regular spatial patterns from an initial uniform state (Turing, 1952). More recently, many researchers have developed computational simulations that employ state-of-the-art knowledge about the properties and interactions of individual molecular components to attempt spatially-resolved reproduction of complex cell biological phenomena. Reaction-diffusion models of intracellular biochemistry have been used to explore a variety of cellular symmetry-breaking processes, including the establishment of cell polarity (Jilkine and Edelstein-Keshet, 2011). Although the exact mechanistic details vary, spatial simulation approaches have yielded insights for symmetry-breaking systems as diverse as yeast bud site selection (Wedlich-Soldner et al., 2003), asymmetric cell division in early *C. elegans* embryos (Dawes and Munro, 2011) and neutrophil chemotaxis (Onsum and Rao, 2007). Establishment of spatial gradients that determine cell fate has been explored in cells ranging from giant syncytial *Drosophila* embryos (Gregor et al., 2007) to tiny individual bacteria (Chen et al., 2011). Simulations that explicitly consider spatial effects as one cell communicates with its neighbors have been used to understand the formation of regular stripes in *Drosophila* embryos (von Dassow et al., 2000), and bizarre non-cell-autonomous effects in the patterning of wing bristles in adult flies (Amonlirdviman et al., 2005).

In many of the modeling efforts that we have mentioned so far, the computational model and/or simulation was formulated explicitly for the problem at hand. While this approach has enabled important scientific insights, we believe that spatial simulation for cell biological processes can become a much more widely used tool in the cell biologists' toolbox if there were more general access to user-friendly implementations of general spatial modeling frameworks that do not require extensive computational expertise to use. Consider, for example, the wide variety of user-friendly open-source software packages now available for analysis of sequencing data (Rice et al., 2000; Trapnell et al., 2012). One particularly important benefit of more standardized approaches to spatial simulations in cell biology is that standardization may help to resolve whether conflicting conclusions arise because of fundamental scientific differences in model assumptions or because of details of numerical implementation.

Another aspect gaining importance as we zoom in closer on the building blocks of cellular structures is the fact that, at the molecular level, cellular biochemistry is governed by stochastic processes such as thermal Brownian motion and the collisions and interactions among individual particles (Schnoerr et al., 2017). Only in

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the limit of high concentrations and homogenous spatial distributions can the behavior of the molecular components of cellular signaling pathways be described in terms of deterministic reaction rate equations. Many subcellular mechanisms operate far from this limit either due to highly non-homogenous clustering of receptors and of the signaling components they recruit, such as studied in the MAPK signaling pathway (Takahashi et al., 2010), or due to locally low copy numbers, for instance of multi-molecular complexes regulating transcription in the nucleus (Cho et al., 2018). To accurately capture the stochastic characteristics of such processes, computational models have to simulate the motion and interactions of individual molecules. A full simulation of Brownian dynamics, following each single 'Brownian hop' of all molecules of a cellular region would, however, in most cases be too computationally expensive and time-consuming. Moreover, by choosing a particular time step for recreating Brownian hops on the computer, we would impose this time scale on our simulations, missing events, such as molecular encounters, that may occur 'in between' our time steps. Many approaches have been developed to deal with this problem and we will discuss several of them below. A common theme among all of them is, however, saving computational cost through temporal and spatial coarse graining without sacrificing too much accuracy in the simulation results. This challenge unites researchers looking at spatially-resolved modeling (without a focus on stochastic effects) and those that try to capture the manifestations of stochastic fluctuations in cellular systems.

The question of how the field could best go about building and sharing broadly applicable computational tools for spatial and stochastic modeling of cell biological processes was the focus of our working group, organized by J.R.F. and R.F.M., which met with the support of the National Institute for Mathematical and Biological Synthesis (NIMBioS). We chose to focus specifically on the biochemical scale of molecular interactions. We did not attempt to include the enormous field of molecular dynamics (MD) simulations that explores forces and movements of individual atoms within proteins or other macromolecules. The reason is that, because of their intensive computational demands, MD simulations are currently limited to exploration of very small biological systems (a few macromolecules) over very short periods of time (typically in the sub-microsecond range), too small and too fast to be incorporated into cell-scale computational simulations. Conversely, we also limited ourselves to considering simulations of interactions within systems of molecular complexes with specific stoichiometry rather than extending our analysis to mesoscopic-scale models that abstract the behaviors of complex molecular systems into continuum physical descriptions, such as those describing cytoskeletal filaments as elastic beams (Nedelec and Foethke, 2007a; Odell and Foe, 2008) or the plasma membrane as a flexible thin film (Fowler et al., 2016). While these kinds of models are enormously useful in cell biology, they rely on fundamental simplifying assumptions. In contrast, we are specifically interested in exploring whether detailed simulations of the behaviors and interactions among bio-molecular complexes can succeed in predicting certain kinds of mesoscopic phenomena, and hence may help determining under which conditions the simplifying assumptions are justified. As we will discuss in our conclusion, it will be an exciting future direction for the field of biological simulation when all these three levels of spatially-resolved simulations can be seamlessly interconnected.

In this article, we will first briefly survey several existing approaches for spatial and stochastic cell simulations and then apply them to a series of 'unit tests' and benchmark problems. The problems we chose are categorized to cover a variety of different aspects of spatially resolved and/or stochastic simulations of cellular behavior. Conceptionally, they are relatively simple and are meant to capture specific challenges related to essential aspects of biological processes. They must be accurately modeled by a simulation tool to ensure that that tool's results are reproducible for the problem category covered by the unit tests or benchmark problems.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

With these examples in hand, we then summarize how features of a particular cell biological problem should guide selection of the appropriate modeling approach. Finally, we will present the results of our wide-ranging and, sometimes, highly opinionated discussions on future directions and challenges for the field. We all share an ambitious vision of the future power of spatial cell simulations both for exploring hypotheses about mechanisms and for coming to grips with the massive amounts of quantitative data now available to cell biologists. Our overall goal here is to map out where the field currently stands and propose a trajectory for the future.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

MODELING APPROACHES

Overview of approaches for spatial and stochastic modeling and simulation of molecular reactions underlying cell-biological phenomena

Computational models that simulate the biochemistry underlying cell-biological processes need to be able to describe molecular players and their reactions. However, depending on the particular question at hand, taking into account spatial aspects and stochastic effects (Fig 1) may or may not be essential, as we show in the Results section. The addition of spatial resolution is computationally demanding, and stochastic simulations are usually more costly than their deterministic counterparts. To attempt a whole-cell simulation, for example, one must choose whether more components and a more complex reaction network is necessary, generally requiring the sacrifice of spatial resolution (Sanghvi et al., 2013) or if spatial resolution is necessary, then the reaction network must be simplified (Ghaemi et al., 2020). If spatial resolution is a priority, must species be resolved as individual particles, capturing fluctuations in copy numbers but at considerable extra expense (Table S3), or is an efficient deterministic approach sufficient?

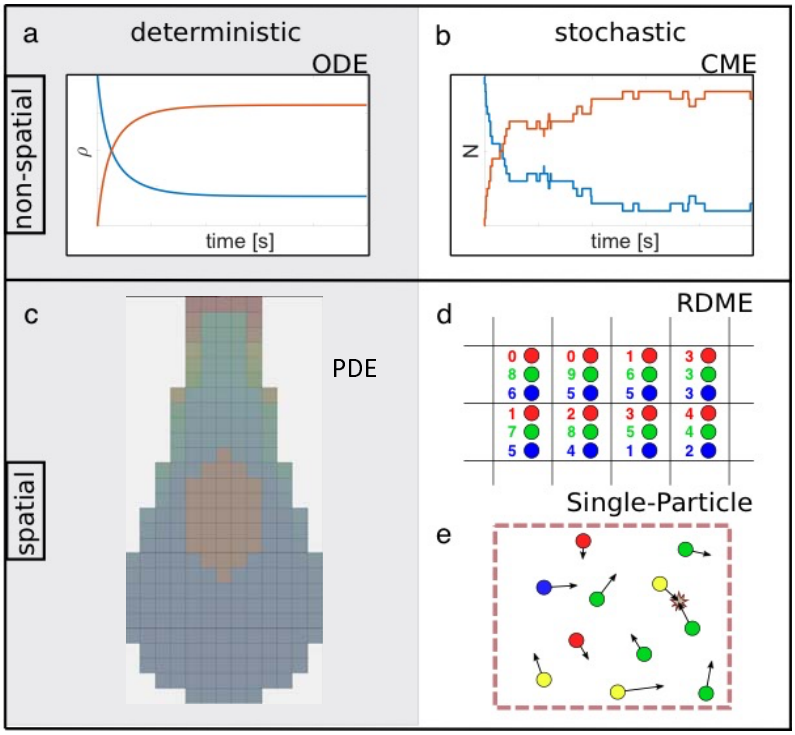


Fig. 1: Overview of non-spatial and spatial simulation approaches for describing the time-dependence of interacting and reacting species. Different colors represent different species. Spatial models are illustrated at one point in time. Left: Deterministic approaches to modeling biological systems solve a) ordinary or c) partial differential equations (ODE/PDE) using standard numerical methods, benefiting from extensive method development across science, math and engineering fields. PDEs are numerically solved on a mesh as shown. Right: Stochastic simulation approaches sample from a time-(and space-) dependent probability distribution that typically models unimolecular and bimolecular reactions, as well as diffusion for spatial methods. d) Reaction-Diffusion Master Equation (RDME) methods are the extension of the b) Chemical Master Equation (CME) methods on to a spatial lattice, where integer copy numbers of species are tracked and can diffuse between lattice subvolumes. e) Single-particle methods propagate individual particles undergoing diffusion in continuous space, where bimolecular reactions can occur only upon collision or co-localization in space. We provide a guide to corresponding software tools of each approach in Table S1.

Non-spatial modeling approaches

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

In many situations, we can describe a biochemical system adequately in terms of the overall concentrations of interacting molecule types and complexes (collectively called 'species'), while neglecting the spatial variations in these concentrations. Reaction rate equations (**see Box 1**) describe how the species concentrations evolve in time. The terms in these equations arise from the rates of the reactions that can occur in the system, which are often described by the Law of Mass Action. The rate of reaction between two interacting species can be given by the product of their concentrations and a rate constant, e.g., k_{on} for the ligand-receptor binding described in Box 1. The bimolecular rate constants that appear in these equations are sometimes referred to as 'macroscopic' rate constants because they describe the average rate of reaction assuming homogenous distribution of the reacting species. In contrast, 'microscopic' rate constants, which govern reaction kinetics at the scale of interacting particles, may take into account more details about the way the molecules approach each other, as discussed below.

Box 1: Reaction Rate Equations:

Technically speaking, reaction rate equations are ordinary differential equations (ODEs). Here, the 'ordinary' refers to the fact that they involve only time (as opposed to, for instance, time *and* space). To describe the time evolution of multiple interacting molecule types, one uses coupled differential equations that express how the components' concentration changes are linked (or coupled). For applications, see, for example (Aldridge et al., 2006; Tyson et al., 2003). From a numerical/mathematical point of view, ODEs describing biochemical reactions are typically simple and many tools exist that can solve them to obtain the temporal evolution of the concentrations in ODE models.

Consider a simple model of a receptor binding to a ligand. We call R the concentration of the receptor, L that of the ligand, and RL that of the complex formed by the binding of the two. The rate equations giving the time derivatives of RL, R and L for this reaction could be written as

$$\frac{dRL}{dt} = k_{\mathrm{on}} \cdot R \cdot L - k_{\mathrm{off}} \cdot RL$$
$$\frac{dR}{dt} = -k_{\mathrm{on}} \cdot R \cdot L + k_{\mathrm{off}} \cdot RL$$
$$\frac{dL}{dt} = -k_{\mathrm{on}} \cdot R \cdot L + k_{\mathrm{off}} \cdot RL$$

Here, k_{on} and k_{off} are the association and dissociation constants, respectively. The time course of RL would look similar to the red curve in ***Fig.1a***, whereas time courses of R and L would be similar to the blue curve. These equations can be solved analytically, but the additional complexity of most biologically-relevant models generates equations that require numerical solution by computer.

The chemical master equation considers the discrete and stochastic nature of the biochemical system. The CME describes how the probability of the system being in a specific state evolves over time, by using reaction probabilities (likelihood of occurrence per unit time) rather than the equivalent reaction rates. Just like reaction rate equations, the CME assumes well-stirred (homogeneous) systems. In most practical modeling applications the CME cannot be solved analytically, but simulations of the CME are conceptually straightforward and very widely used (**see Box 2**).

Both the reaction rate approach and the CME approach simply cannot capture effects of inhomogeneous distributions of molecules in space, such as receptors clustered on membranes, or intrinsic time delays due to diffusion to localized targets such as membrane bound receptors. For more realistic

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

simulations of cellular process, we must turn to different computational approaches that explicitly include space.

Box 2: **The Chemical Master Equation (CME):**

The chemical master equation (CME) is a set of coupled linear ODEs that describe the time-dependence of the probability to occupy each of a set of discrete states, with each state defined by the copy numbers of molecular species. The CME may be written as

$$\frac{dP(N,t)}{dt} = \sum_{R} a_r(N - v_r) \cdot P(N - v_r, t) - \sum_{R} a_r(N) \cdot P(N,t).$$

where the composition vector N is composed of copy numbers of each molecular species, and thus one has a set of equations, one for each possible instantiation of *N*. Here, the sum runs over all possible reactions. v_r is the stoichiometric vector of reaction r that describes how this reaction changes the number of molecules in the composition vector N and $a_r(N)$ is the probability per unit time that reaction r occurs, given that the system is in the state described by N. The CME is important from a conceptual point of view as it represents a framework to describe probabilistic transitions and thus captures the stochasticity underlying all molecular interactions(Schnoerr et al., 2017) (Grima and Schnell, 2008). The computational cost of solving the CME equations scales exponentially with the number of chemical species, and, although clever approaches have extended the size of systems for which the CME can be solved (Munsky and Khammash, 2006), high computational cost still limits biological applications. An intuitively simple way to calculate a solution of the CME would be to set up a simulation where time ticks forward in small, discrete intervals (time steps). However, the fixed time step would always introduce the potential for error since reactions may occur even during shorter time steps than the one chosen to propagate the system in time.

One popular and precise method used to generate trajectories through the state space sampled by the CME without the need to choose a discrete time step is the Gillespie algorithm, also known as the stochastic simulation algorithm (SSA) (Bortz et al., 1975) (Gillespie, 1976a). In a Gillespie simulation, the time interval until the next reaction occurs is itself sampled, as is the type of reaction that will occur (Gillespie, 1976a). Molecular species that can react quickly and have many possible interaction partners will be selected frequently while rarer molecules associated with slower reactions will be selected rarely. Since the simulations proceed with one reaction at a time, the computational cost depends strongly on the number of particles and reaction rates in the system. In contrast, the effort required to integrate (or solve) reaction rate equations depends mostly on how many molecule types are involved and whether their interactions occur on different or similar timescales. Various approaches have been developed to increase the efficiency of both exact (Gupta and Mendes, 2018) and approximate (Schnoerr et al., 2017) stochastic simulations of the CME. In addition, efficient methods have been developed to compute distributions and moments directly from the CME itself (Hasenauer et al., 2014). A stochastic simulation of the example system from Box 1 would look similar to ***Fig. 1b***. Note, however, that such a trajectory represents only one possible time course compatible with the underlying CME. This means that many stochastic simulation trajectories must be collected to determine probability distributions and moments of the CME.

Spatial modeling approaches

The molecular components of living systems are not distributed homogeneously and the high spatial resolution of today's fluorescence microscopy is continuously giving us more examples of biological

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

phenomena where the spatial arrangement of the underlying biochemical processes is fundamentally important. To model such phenomena we have to switch from non-spatial to spatial simulations. However, this switch is frequently not easy due to a growth in the number of model and system features that must be specified (Fig. 2a). The most important difference between non-spatial and spatial simulations is that the latter take into account the translocation of the interacting molecules. In the simplest case, this means that, in addition to reactions, the diffusion of molecular species in space must be simulated. Along with diffusion, the system geometry must be specified and what happens at the 'walls'. This is significantly more challenging to implement when the spatial system containing the interacting molecules is not simply a square box with rigid walls. Realistic spatially resolved models often aim to capture aspects such as particular cell morphologies; examples include synaptic structures with narrow regions connected to larger cell bodies (Cugno et al., 2019; Rangamani et al., 2016) or geometries that are flat and almost two-dimensional, like lamellipodial extensions in migrating eukaryotic cells (Nickaeen et al., 2017) (Fig. 2b). Geometries for cell simulations can be designed by hand, derived from microscope images (Schaff et al., 2000), or generated from machine-learned cell models (Majarian et al., 2019). We note that spatial models have the capacity to build in many additional features such as mechanics, electrostatics or coarse molecular structure that may be particularly important for useful simulations of some cell biological processes (Fig. 2d). There are very few existing computational tools that take these additional features into account, largely because they introduce further challenges to both the numerical implementation and the mathematical descriptions of the physical model.

Reaction-diffusion equations represent the most straightforward extension of reaction rate equations for the inclusion of spatial aspects. Instead of depending just on time as a variable, the behavior of molecular species now additionally depends on spatial coordinates. Equations describing these reactions must therefore be formulated as partial differential equations (PDEs) rather than as ordinary differential equations (ODEs) (Fig 1c). As is frequently the case for mathematical models of biological phenomena, only very simple situations can be described through equations that can be explicitly solved in such a way that the solution describes the behavior of the modeled system as a continuous function of space and time (Lipkow and Odde, 2008). In most cases, one has to explore reaction-diffusion equations through numerical simulations that divide space into sub-volumes (frequently called 'voxels') and calculate how diffusion leads to exchange among the sub-volumes. Reaction-diffusion equations have been widely used to model spatially-resolved biomolecular dynamics and interactions of cell-biological systems (Loew and Schaff, 2001). The spatial dynamics can be extended beyond pure diffusion (e.g. to include advection), and reactions can be defined phenomenologically (Hill-type or Michaelis-Menten). Like reaction rate equations, deterministic PDE reaction-diffusion equations do not capture stochastic fluctuations in species numbers, and thus cannot, for example, capture pattern formation driven by a system's sensitivity to low copy numbers (Howard and Rutenberg, 2003).

One way that stochastic reaction-diffusion equations can be formulated is the spatial extension of the chemical master equation, known as the Reaction Diffusion Master Equation (RDME) (Fig 1d). Instead of only defining how a system switches from one set of numbers of molecules in particular states to another (for instance when a complex in the system decays into two molecules) the RDME includes 'hops' from one location to another. Importantly, just like the CME, the RDME describes discrete changes. That means, it requires a spatial discretization into sub-volumes, within each of which well-mixed conditions are assumed to prevail. Diffusion events of molecules are tracked only when they occur between adjacent sub-volumes, not within an individual sub-volume (Fange et al., 2010). Similar to the non-spatial case, it is usually not possible to solve the RDME analytically, and instead it is standard practice to compute solutions by simulating a particular

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

stochastic time evolution of the system with, for instance, a Gillespie approach (see Box 2) that adds diffusional hops to the list of events that can occur. However, care must be taken to choose the right degree of spatial resolution that strikes the appropriate balance between capturing spatial details and avoiding sub-volume sizes that are so small that discretization dilutes the molecules to a point where they essentially do not 'see' their potential reaction partners anymore because the molecules are spread out over distinct sub-volumes.(**see Box 3**).

Box 3: Spatial Discretization of Reaction Diffusion Equations:

To numerically solve reaction-diffusion processes modeled as partial differential equations (PDEs), it can be challenging to choose the appropriate spatial discretization of the modeled biological geometry. A discretization that is too coarse will suppress many spatial details and will represent a poor approximation of the underlying biology. However, keeping track of the contents of many very small voxels will not only be very expensive computationally, but may also lead to situations where the assumptions of mass-action kinetics no longer strictly hold since the fraction of the molecules in the system that populate a single voxel becomes so small that the very concept of an average concentration becomes problematic. Furthermore, the accuracy and efficiency of PDE solvers is not just sensitive to the resolution of the spatial discretization (sometimes called lattice or mesh) but also to the discretization scheme as manifested, for instance, in the shape of the voxels. The practice of designing adaptive meshes, that is combining voxels of different shape and size in one simulation to capture small-scale spatial details where needed while keeping the total number of voxels as low as possible, is a field of active research. The structure of the mesh also has to be adjusted to the numerical method chosen to solve the reaction-diffusion PDEs. Finite volume methods directly simulate diffusional exchange between voxels. In contrast, finite element algorithms optimize the coefficients of interpolation functions at the nodes of the mesh to achieve good approximations of concentration profile resulting from the combination of reactions and diffusion. See, for example (Richmond et al., 2005).

Similar to PDEs, the accuracy and cost of the ***Reaction Diffusion Master Equation*** (RDME) is sensitive to the spatial mesh; this problem is inherent to all spatially discretized simulations. Computational costs grow rapidly as the mesh resolution increases. Importantly, the accuracy of an RDME model does not always increase with a finer mesh. A very small mesh size violates the assumption that species are dilute and their own molecular volume is small relative to the voxel (Erban and Chapman, 2009) (Isaacson and Zhang, 2018; Wolf et al., 2010) . For specific non-fundamental reaction types, RDME has an additional limitation in that it does not always converge to the CME solutions in the limit of fast diffusion, as expected (Smith and Grima, 2016). Hence, the RDME may be viewed as a *non-converge*nt approximation of more microscopic spatially continuous models, such as the Smoluchowski model to be discussed below.

Particle-based spatial simulation methods take into account the stochastic motion and interactions of individual molecules in continuous time and space and are thus capable of modeling biochemical processes that involve low copy numbers and strongly heterogeneous molecular spatial distributions (Fig. 1e). These methods have the highest resolution (Fig. 2c) but they come with a high computational cost. Importantly, simulations that treat each particle as an individual also offer the possibility of building in more detailed molecular features (Fig. 2d). Typically, particle-based approaches resolve a bimolecular reaction A+B→ C of a pair of molecules as two physically distinct stochastic processes. First, the molecules' diffusive (Brownian)

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

motion leads to their encounter. Then, the molecules either form a bond with a reaction probability determined by the reaction rate constant and their current separation, or else they diffuse away from each other. Numerical approaches to single-particle reaction-diffusion calculations for biological systems can be usefully grouped into two classes; first, what we will call Smoluchowski-based methods, and second, what we will call single-particle Kinetic Monte Carlo (KMC) schemes (Fig. 2c).

The definition of the reaction probability is the primary challenge and distinguishing feature of different single-particle algorithms. The first class of single-particle methods we will consider, Smoluchowski-based methods, derive reaction probabilities from the Smoluchowski model of diffusion-influenced reactions (Collins and Kimball, 1949; von Smoluchowski, 1917). These methods explicitly recognize that the macroscopic kinetics of association must depend on both diffusion and reaction rate constants, resulting in a distinction between a microscopic and macroscopic rate (**see Box 4**).

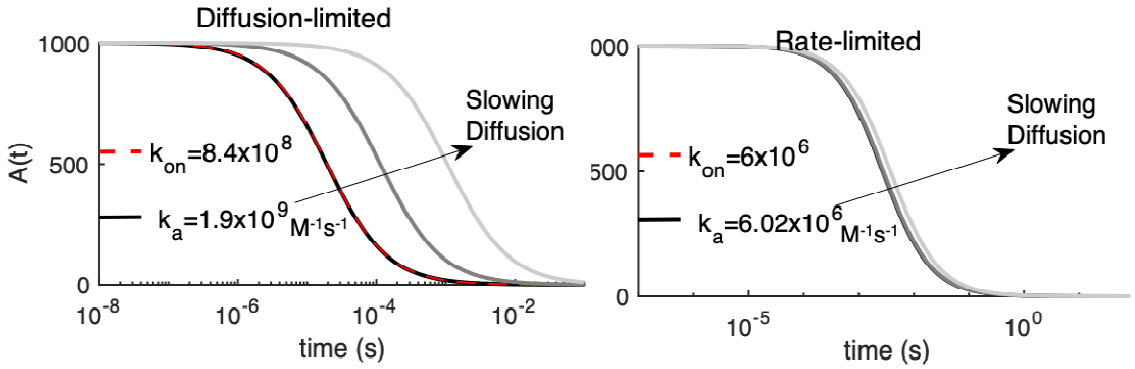
bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

BOX 4: Microscopic vs Macroscopic Rates, and the Sensitivity of Strong Binding to Diffusion:

For all non-spatial models, as well as PDE, RDME and KMC-type spatial models, bimolecular association reactions are parameterized by the macroscopic rate constants, k_{on}, corresponding to the rates one would measure from a binding experiment in bulk solution. This is because in all these spatial models, species that are localized in a small volume are assumed well-mixed, thus obeying the same mass-action kinetics used in non-spatial models. In the Smoluchowski-based single-particle methods, however, molecular interaction kinetics are split into two steps, as described in the text. This results in a purely diffusive contribution to the bimolecular collision, and then what is effectively a purely energetic contribution defined by a microscopic on-rate k_a, assumed to occur at a specific binding radius . In 3D, this results in the long-known relationship:

— —— , Eq. A

where *D* is the sum of both species diffusion coefficients. With this relationship we can directly assess the impact of diffusion on controlling macroscopic kinetics. As shown in the Figure, for large microscopic rates, the macroscopic kinetics of the A+B→∅ reaction is noticeably dependent on diffusion, whereas for smaller rates, the effect of diffusion is negligible, despite D_A=D_B dropping from 100 to 1 um^2/s. Here A_0=B_0=62 M. Hence large k_a, or strong binding, is diffusion-limited, and small k_a is rate-limited.



In two and one dimensions (e.g. on surfaces and filaments), the relation between microscopic and macroscopic parameters is more complicated, due to the properties of diffusion in lower dimensions smaller than three. No single relationship exists between microscopic and macroscopic rates [see e.g. (Yogurtcu and Johnson, 2015)], but meaningful theoretical relationships can be defined if the system size is considered (Szabo et al., 1980), where in 2D we further correct for system density using (Yogurtcu and Johnson, 2015): — — ——— ——— , Eq. B

where ——— .

and are the copy numbers of reactants A,B and the system size is S.

For our test cases below, we thus always derive the microscopic rates k_a to reproduce the macroscopic rates using Eq A or Eq B. This is because it is already clear from non-spatial models that changes to the macroscopic rates will necessarily alter the reaction kinetics. Since we are not focused on probing the influence of kinetic parameters on molecular behavior, but rather the role of explicit spatial representations in controlling species distributions and encounter times, we preserve all k_{on} values. However, it is worth noting that for a reaction pair with a large k_a value, if diffusion slows throughout the simulation due to, for instance, formation of large complexes, then the macroscopic kinetics will also slow down, an effect that is naturally captured in Smoluchowski-based methods.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Calculating the positions and reactions of all molecules in this framework by “brute force” typically requires computationally costly simulations using extremely small time steps to minimize the risk of missing molecular encounters. An alternative approach uses a mathematical formalism that predicts the encounter probability for pairs of particles based on the particles’ initial positions. This ‘Green’s function’ approach can, however, only calculate the encounter probability for two particles at a time. This means that the simulated system has to be segmented into two-particle subsystems (**see Box 5**). In practice, this turns out to be feasible for many interesting biological problems. Several computational frrameworks have been developed to take advantage of this approach, including FPR (Johnson and Hummer, 2014), NERDSS (Varga et al., 2020), GFRD (van Zon and ten Wolde, 2005) and eGFRD (Sokolowski et al., 2019) (Fig. 2c). A related computational implementation, known as Smoldyn, approximates the solution to the Smoluchowski model (Andrews, 2017; Andrews et al., 2010; Andrews and Bray, 2004). By approximating the solution, Smoldyn is simpler to implement than the more strictly constrained implementations listed above, but the time-dependence and in some cases equilibrium are not as rigorously correct.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Box 5. Single-particle reaction-diffusion simulations:

Single-particle methods simulate the stochastic behavior of individual and (in spite of the name) pairs of interacting molecules. Any biochemical network whose description does not include ad-hoc phenomenological processes (such as, for example, Hill coefficients describing non-linear dose-response characteristics) can be described as composed of uni- and bimolecular reactions. As unimolecular reactions are only time-dependent, they are typically modeled as Poisson processes. For bimolecular reactions, the distance between a pair of particles influences the probability that they will diffuse to collision and react with one another in a time-step. The time evolution of the molecules' positions is described by a stochastic differential equation, the overdamped Langevin equation(Van Kampen, 2007). Its numerical implementation, known as Brownian Dynamics (BD) (Ermak and Mccammon, 1978; Northrup et al., 1984), requires tiny time steps to accurately resolve molecular encounters, which renders the BD scheme highly inefficient. It is, however, possible to introduce a shortcut by using Green's function approaches to calculate the distance-dependent reaction probabilities for pairs of molecules. This can enhance the efficiency of BD simulations by resolving bimolecular reactions within one large time step, Δt.

For the Smoluchowski model, the Green's function (GF) $p(r, \Delta t \,|r_0)$ can be obtained as the solution of the diffusion equation that describes a pair of molecules A, B that diffuse with diffusion constants D_A, D_B, respectively, and may undergo a reaction $A + B \rightarrow C$ as follows

$$\frac{\partial p(r, t|r_0)}{\partial t} = D\nabla^2 p(r, t|r_0)$$

where $D = D_A + D_B$ and r, r_0 refer to the distance between A and B after and before the time step, respectively. In accordance with the two-step picture described in the main text, reactions are incorporated by imposing boundary conditions that specify the physics at the encounter distance $r = \sigma$, such as Smoluchowski-Collins-Kimball BC (Collins and Kimball, 1949)

$$4\pi\sigma^2 D \frac{\partial p(r, t|r_0)}{\partial r}|_{r=\sigma} = k_a p(r = \sigma, t|r_0)$$

where k_a refers to the intrinsic reaction constant. It is worth noting that finding the appropriate Green's function to capture the desired properties of the molecular interactions can be a challenge. For the case of *reversible* diffusion-influenced bimolecular reactions of an isolated molecule pair in 2 dimensions the GF was derived only in 2012 (Prustel and Meier-Schellersheim, 2012).

The second class of particle-based approaches we will consider, single-particle Kinetic Monte-Carlo (sp-KMC) schemes, are also based on the two-step picture of bimolecular reactions. However, the sp-KMC methods do not seek to strictly implement an underlying physical model, unlike the Smoluchowski-based models. Instead, the reaction probabilities are derived by using various approximating assumptions. Widely-used implementations include MCell (Kerr et al., 2008), ReaDDy (Schoneberg and Noe, 2013), and SpringSaLaD (Michalski and Loew, 2016). Within these frameworks, the mathematical expressions of reaction probabilities typically assume a simpler, computationally more efficient form than their Smoluchowski model counterparts. While this is obviously useful for improving computational speed, it may not be immediately clear to what extent the approximations are justified and how much error they introduce. While both classes of model can predict the same behavior in simple systems (Tauber et al., 2005), Smoluchowski-based algorithms inherently capture excluded volume (particles exclude a sphere of radius σ relative to partners) and microscopic collisions, whereas sp-KMC methods do not. Smoluchowski models thus provide higher resolution at short distances and for dense systems.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The future extensibility of both classes of approach hinges on the feasibility of finding mathematical expressions for the crucial reaction probabilities that incorporate additional features and details, such as curved surfaces, intra-molecular constraints, and external and internal deterministic forces (Fig 2b and 2d). Arguably, this task is easier for sp-KMC methods, because they rely on approximations rather than exact analytical solutions to derive these expressions. However, as discussed, this flexibility may come at the expense of a clear physical justification of the underlying assumptions and rigorous error control.

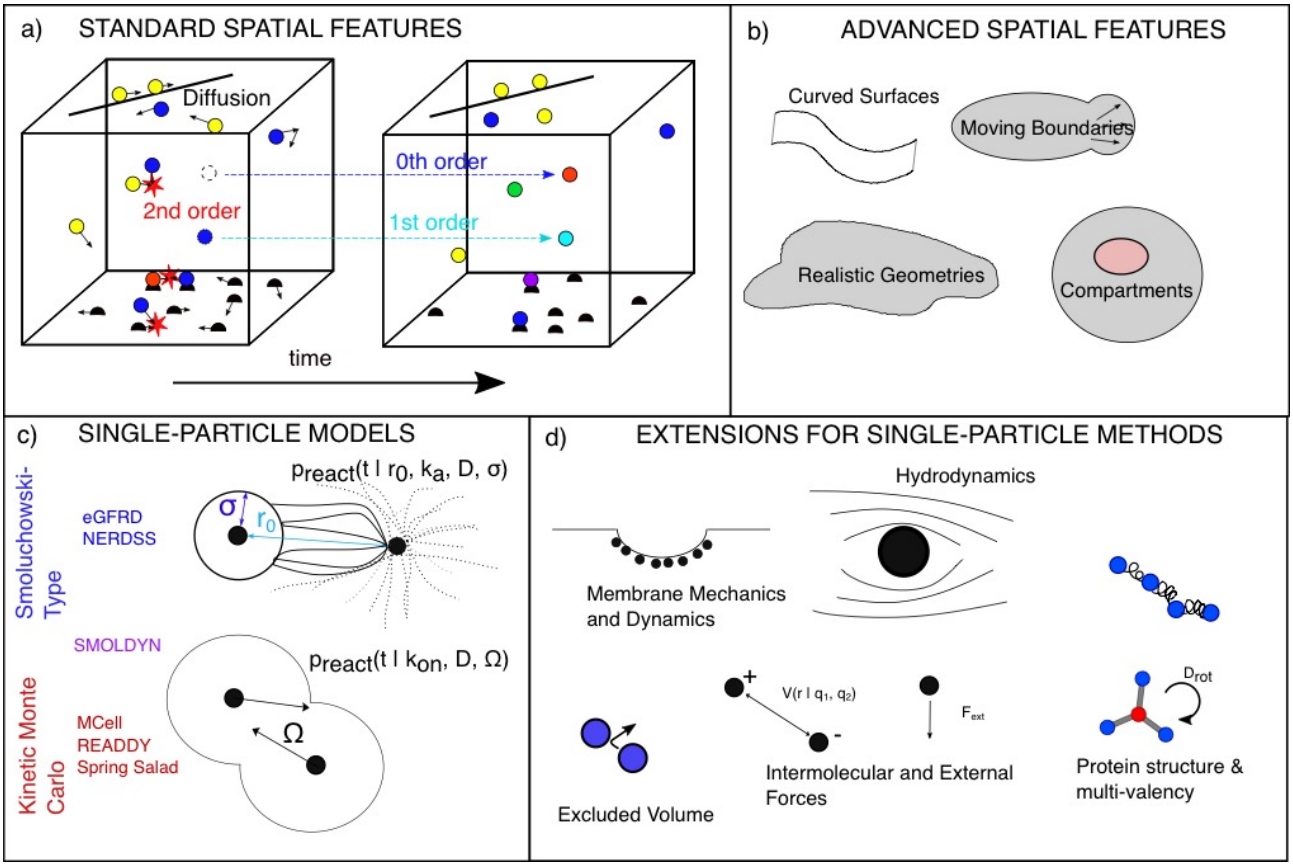


FIGURE 2: **Demands on spatial approaches and their extensibility.** a) For all spatial models, in addition to treating fundamental reaction types in all dimensions (3D, 2D,1D) and between all dimensions (for instance molecular exchange between bulk (3D) and surfaces (2D) they must specify an equation of motion, typically diffusion, and boundary conditions on the system geometry. b) More advanced treatments of curved and complex boundaries require additional care for treating reactions and diffusion. c) Single-particle spatial methods fall into two broad classes. In Smoluchowski-type methods, reaction probabilities (p_{react}) are determined by finding analytical solutions for the fraction of diffusive trajectories that produce collisions that lead to reactions (not all collisions lead to reactions when the microscopic rate $k_a<\infty$). In single-particle Kinetic Monte Carlo methods, algorithmic variations exist but particles that are in a neighborhood Ω can react according to a macroscopic rate, k_{on}. Smoldyn, for instance, is an approximation to the Smoluchowski model. d) Single-particle methods have the capacity to build in higher-resolution features, although these will alter their equation of motion, requiring new definitions of p_{react}. We note that PDE-based models can also expand

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

beyond purely diffusive dynamics. In Table S2 we summarize features available in commonly used software tools.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

RESULTS

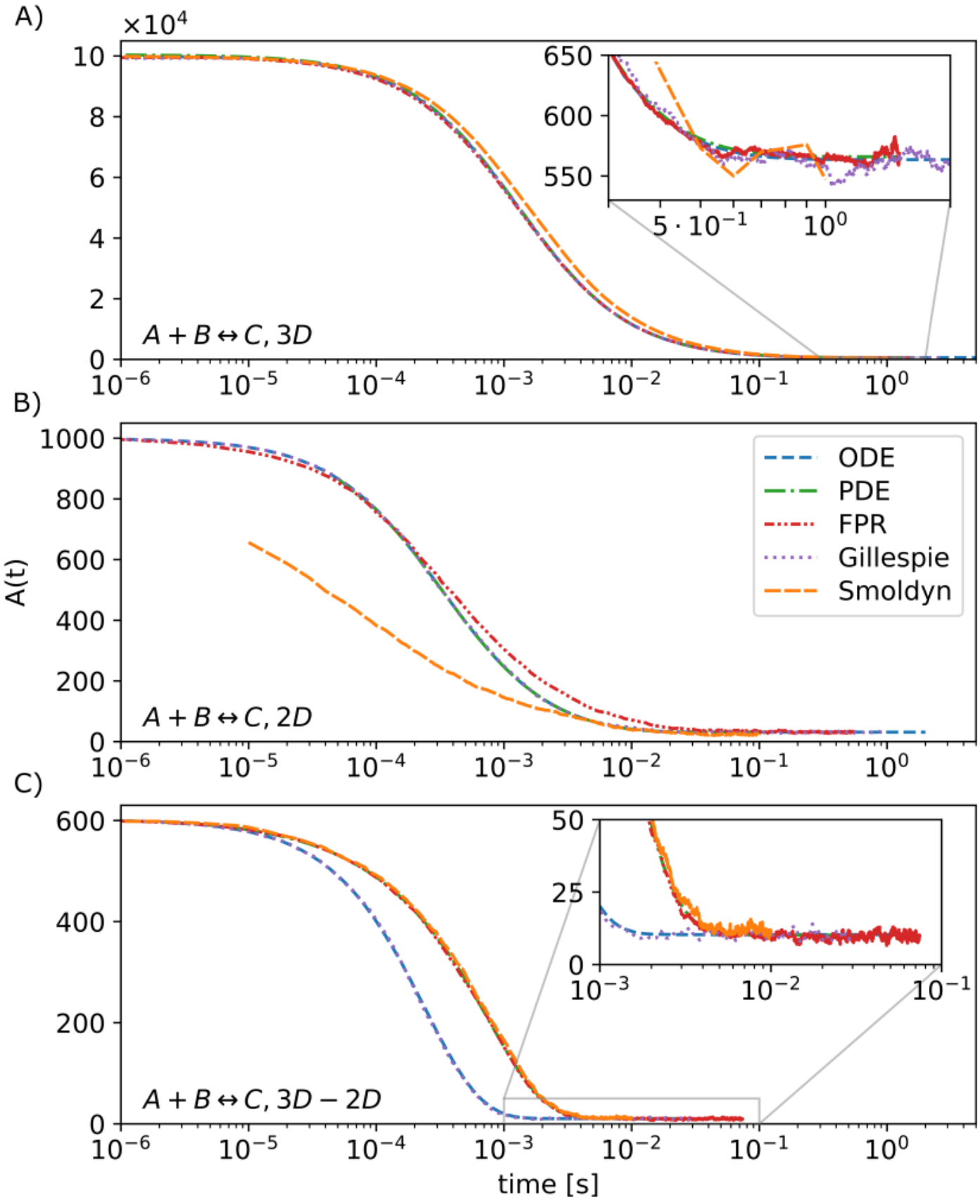
We present a series of test problems relevant to spatial modeling of cellular and subcellular processes. This list is not meant to be exhaustive, but rather to permit a manageably-sized survey of the kinds of problems that different simulation programs may be challenged to solve within a larger biological study. Our selection thus includes very simple problems that can be solved by many simulation tools as well as complex problems that can be solved (at present) by only a few. By framing the problems explicitly, we facilitate a direct comparison among different simulation packages both with respect to accuracy of execution and how they encode these particular scenarios. While all the models presented have been simulated previously, they have not been subject to the quantitative comparative analysis performed here across multiple model and method types. This comparison provides us with distinctive insight into the sensitivity of quantitative and qualitative behavior that emerges with specific biologically relevant features, which we summarize at the end of the Results section.

For these test cases, we contrast results from stochastic, deterministic, spatial and non-spatial modeling approaches. We use Virtual Cell software for all non-spatial simulations and spatial deterministic simulations. For spatial stochastic simulations, we use the single-particle softwares NERDSS, Smoldyn, MCell, and eGFRD. We provide executable model inputs and numerical outputs for each model in our publicly accessible repository (https://github.com/spatialnimbios/testcases/). We summarize a broader range of actively developed tools and their features in Table S2, as distinct software tools have introduced selected complex features of RD systems (also discussed previously (Schoneberg et al., 2014; Takahashi et al., 2005)), and this is an additional consideration for users when selecting a tool for their biological problem.

Category 1: "Unit test" cases

This category of problems represents fundamental building blocks for which there is a known correct answer (at least at steady-state). We emphasize that because these reactions form the basis of much more complex models and geometries, it is essential that they be carefully tested for accuracy with regard to reaction kinetics and, in the case of bimolecular reactions, reversibility. For all of these, we initialize simulations with well-mixed components, and thus one may expect that any modeling approach would give the same outcome. However, due to differences in both approaches and algorithmic choices, we find that differences do in fact emerge in specific parameter regimes, particularly at short times before the system approaches steady state (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3. Reversible bimolecular reactions provide fundamental building blocks for any complex biological systems and thus warrant careful testing. In all dimensions, we use well-mixed initial conditions in closed systems, such that the reactions will reach a well-defined thermodynamic equilibrium. We test them using the non-spatial ODE and Gillespie methods, which match each other for all systems. The analytical solution to the rate-equation is essentially exactly captured by the ODE. We compare the spatial PDE solved using VCell with single-particle methods solved via the FPR algorithm (NERDSS) and Smoldyn. For the 3D→2D reaction, we see the first difference between the ODE and PDE solution, because in the spatial models the reactants are well-mixed in distinct locales (volume vs surface), whereas in the non-spatial models, they are all well-mixed in one volume. For 2D, single-particle methods will differ from other approaches, with NERDSS essentially capturing the theoretical diffusion-influenced reaction dynamics in 2D (Yogurtcu and Johnson, 2015).

1A: Bimolecular association in 3D, 2D, and from 3D to 2D

Although seemingly simple, bimolecular association events require both a diffusional encounter and a reactive event, thus the rate-constants and the kinetics are dependent on the dimensionality of the systems, and even for well-mixed systems, spatial details can cause deviations from non-spatial models. For reversible bimolecular association of well-mixed reactants in a closed system, the equilibrium is theoretically well-defined and the kinetics for non-spatial rate equations can be derived analytically. One may note that when bimolecular association is reversible, recovering the proper equilibrium is a simple test that can nonetheless be challenging for single-particle methods. Reaction probabilities and the placement of reactants upon un/binding events must be derived to ensure equilibrium is reached (Box 5).

In 3D, all models and tools produce nearly identical results, even for this strongly diffusion-influenced reaction (k_{on}=1.48 10^7 $M^{-1}s^{-1}$). This is as expected (Fig. 3a). Although at short times the kinetics is slightly faster for Smoluchowski type simulations (NERDSS), the kinetics rather rapidly converges to the macroscopic rate equations (Johnson and Hummer, 2014). Differences between single-particle and non-spatial methods can also emerge for reversible reactions as they approach equilibrium (Mattis and Glasser, 1998; Tauber et al., 2005), but these can only be effectively observed with high numerical precision and statistics—usually they are dwarfed by the copy number fluctuations.

Unlike in 3D, macroscopic rate equations in 2D only approximate the dynamics captured in Smoluchowski-type approaches at all times (Fange et al., 2010; Hellander et al., 2012; Yogurtcu and Johnson, 2015) (see Box 4). All macroscopic rate-based methods produce the same kinetics as each other (Fig. 3b). Here we see distinctions between the spatial PDE and the spatial single-particle methods. Although species diffuse in the PDE, because they are present at all positions in space (due to uniform initial conditions), association is not dependent on their spatial distribution. For single-particle methods, there is always a distribution of starting separations between species that leads to some very fast reactions initially, and at long times produces slower reactive collisions as particles that started off close to each other have already been consumed in the reaction. Because Smoldyn approximates the dynamics of the Smoluchowski model, the kinetics can be off for specific parameter regimes, with deviations being typically very small in 3D but significant in 2D.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

For binding between 3D particles and 2D particles (relevant for biological cases where soluble cytoplasmic proteins bind to membrane proteins or lipids), all models produce the same equilibrium, but the spatial models have slower kinetics delayed by diffusion to the surface (Fig. 3c). The extent of divergence between the non-spatial and spatial models is driven by three factors, the 'height' the solution volume stretches from the membrane surface, the diffusion coefficient of the 3D particles, and the speed of the binding. Here we simulated a fast, strongly diffusion-influenced reaction (8.4×10^7 $M^{-1}s^{-1}$), meaning nearly every collision results in a reaction. For a simulation box with h=0.2 μm, one can estimate an average time to diffuse to a surface particle would be ~60 μs (D^{3D}=30 $\mu m^2/s$), which is relatively fast. However, with the numerous 2D particles mixed in the solution volume for a non-spatial simulation, the time to diffuse to a 'surface' particle drops to ~6 μs. Thus we find a mean relaxation time of ~200 μs without space, vs ~700 μs with space (Fig 3c). By dropping the reaction rate to more moderate protein-protein interaction levels, the spatial and non-spatial results begin to converge. Smoldyn shows excellent agreement for larger steps, here 10^{-6} s, although the kinetics shift slower for shorter steps. We note that when particles can only collide with one another from one side (because one is embedded in a surface, for example), this reduces the binding by a half, and solvers should explicitly account for this so that user-defined rates produce the equilibrium expected from a non-spatial model.

Lastly, transitions to the surface can be modeled using adsorption, which uses an effectively 1D rate. This is more efficient but importantly, it need not reach the same equilibrium as explicit particle simulations, because the occupancy of surface binding sites is not accounted for. Modeling explicit particles thus gives more control over the surface properties, and algorithms for binding to surfaces while accounting for site occupancy using implicit sites rival adsorption models in speed (Fu et al., 2019). Not all tools allow for all types of surface binding, hence it is important to recognize these distinctions between adsorption vs single-site binding.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

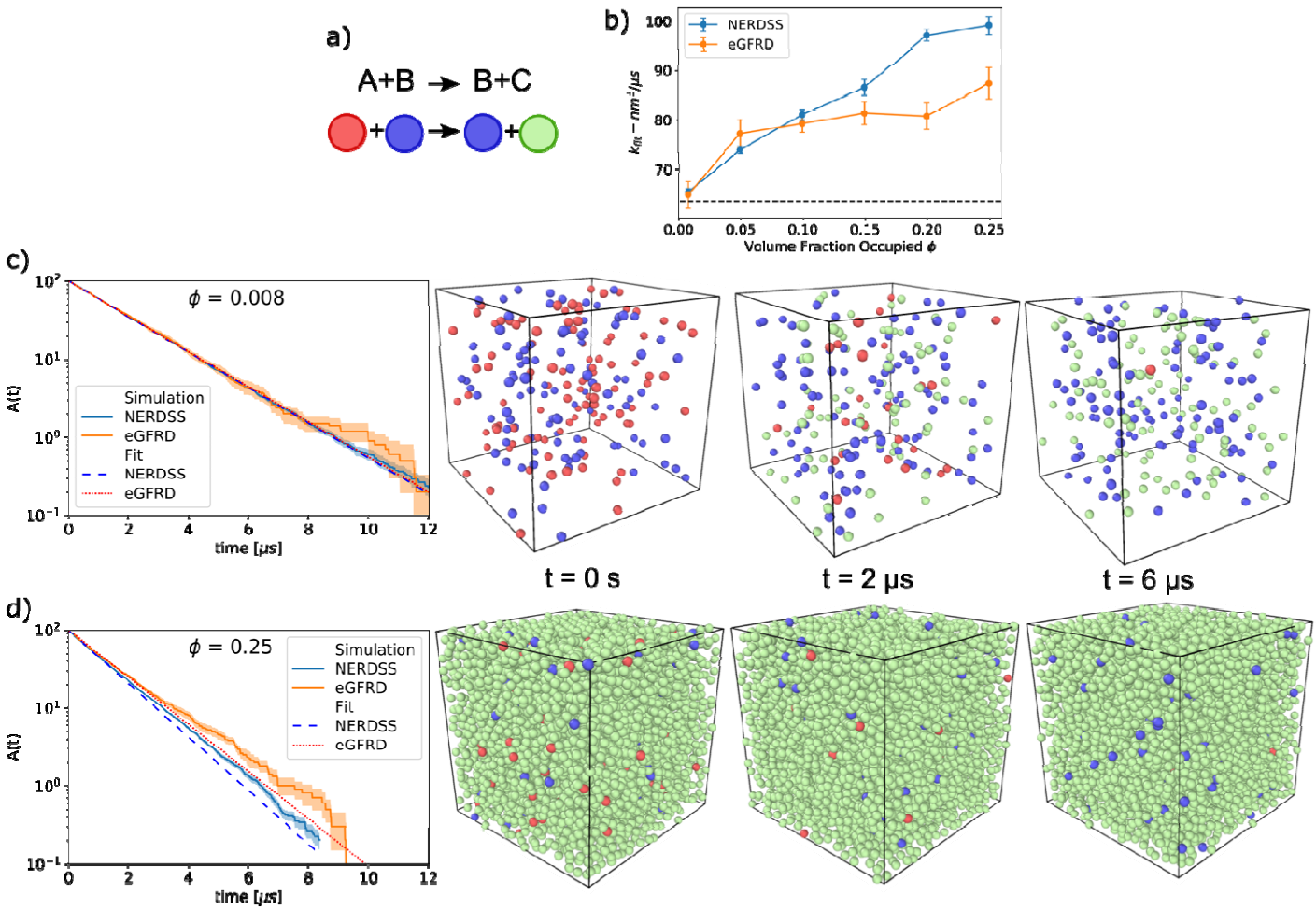


FIGURE 4. The effects of volume exclusion, or crowding, can be tested by a simple model of bimolecular association in the presence of inert particles. a) For the simple model B+A→B+C, the total population of catalytic B molecules (blue) remains fixed. A molecules (red) are converted into C molecules (green) when they collide with B molecules. C molecules do not react, but do exclude volume and so can act as physical crowders. A molecules (red) are depleted over time. b) The kinetics of A depletion depend strongly on the initial concentration of inert crowders C. For each simulation, the rate of depletion of A is fit to a single-exponential based on the solution to the rate-equation (see Methods). The best fit rate is plotted here as a function of the volume fraction occupied by all A, B and C particles. c) For low crowding fraction where the initial concentration of C molecules C_0=0, the kinetics is well described by a single exponential with a rate close to the non-spatial solution of 63.5 (65.4 FPR, 64.9 GFRD). d) For high crowding with a large initial C_0, the simulated kinetics is not as well described by a single exponential, exhibiting slower decay as the density of A approaches zero. Snapshots display the actual volume of the particles (r=0.5nm) given the full volume (23.2nm boxlength), with 100 initial A particles (13.3mM), 100 initial B particles (13.3mM), and variable C. The walls use periodic boundary conditions. Simulations were performed with the NERDSS and eGFRD software, using the FPR and GFRD algorithms, respectively.

1B: Crowding

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Inside of living cells, the extremely high density of macromolecules (with typical spacing on the order of nanometers) can alter the speed of molecular diffusion(Ando and Skolnick, 2010) and kinetics of intermolecular reactions, either increasing or decreasing biochemical reaction rates as compared to rates in dilute solution, depending on the size and mobility of the crowders (Schreiber et al., 2009; Zimmerman and Minton, 1993). As crowders become larger and more immobile (e.g. vesicles) they act more like barriers to encounters, slowing down rates of association (Minton, 2006; Zhou et al., 2008). However, as we see here, when crowders are of comparable sizes and similar mobility to the reactant species, they drive up rates of association. This rate increase is due to a reduction in the total volume available to the reactants, effectively concentrating them without providing substantial barriers to encounters (Minton, 2006; Zhou et al., 2008). To quantify how increasing concentrations of crowding agents alter bimolecular association rates, we here simulated the bimolecular reaction A+B→B+C in the presence of additional, inert, C crowders, where all species diffuse and exclude volume (Fig. 4a). The analytical solution for no crowding/no excluded volume, $A(t) = A_0 \exp(-k_{macro} B_{tot} t)$, provides a convenient baseline and fit function for interpreting deviations due to crowding/excluded volume (Fig. 4b).

For the single-particle algorithms that capture excluded volume (GFRD (van Zon and ten Wolde, 2005) and FPR (Johnson and Hummer, 2014)), two main results emerge. First, the overall kinetics of association increases with increasing crowding fraction, up from k_{on}=63.5 nm^3/μs at zero crowding/no excluded volume (3.8x10^7 M^{-1}s^{-1}) to ~85-100 nm^3/μs with 25% crowding fraction. This result is consistent across both algorithms, indicating that for the well-formulated Smoluchowski model applied to a strongly diffusion-influenced reaction, small mobile crowders will enhance reaction rates for concentrated reactants (here A_0=13.3mM). This same trend was observed for simulations at lower reactant concentrations but comparable rate constants(Kim and Yethiraj, 2009), but differs qualitatively from a recent study which found that crowding reduced reaction rates at the same reactant concentrations studied here(Andrews, 2020). In the most recent study, however, the crowders were completely immobilized(Andrews, 2020), meaning they did not capture thermal motion of macromolecular crowders, but behaved as a rigid and impermeable barrier. Second, for high crowding regimes, we find that the kinetics is not described by a single rate constant (Fig. 4d), whereas at low crowding the results fit extremely well to the non-spatial analytical solution, with a new rate-constant (Fig. 4c). This is perhaps not surprising; for diffusion-influenced reactions, short-timescale kinetics is dominated by reactants that are already close together, where we expect crowding agents to promote their repeated collisions. Then, as the reactant populations decrease either to zero or towards an equilibrium state, the kinetics slows relative to non-spatial predictions, although this shift may be hard to detect (Johnson and Hummer, 2014; Mattis and Glasser, 1998; Tauber et al., 2005; Yogurtcu and Johnson, 2015). Our results suggest that crowding agents exacerbate this slow search for the final reactants, similar to what happens in 2D, making the deviations from a single-rate constant kinetics easier to detect.

Although mobile crowders do impact reaction rates, the changes are often quite modest. For the strongly diffusion-influenced reaction simulated here, we measure clear increases in rates, but for more rate-limited reactions (k_{on}=6x10^5 M^{-1}s^{-1}), the effect of crowders on the reaction rate is minimal (data not shown). Finally, FPR and GFRD are not in perfect agreement in terms of the quantitative size of the change in kinetics, although qualitatively they both predict a higher rate. At these high densities, GFRD converts to a brute-force Brownian Dynamics algorithm, rather than its exact event-driven method. FPR must also take extremely small time-steps (10^{-10} to 10^{-12} s) to prevent particle overlap. It is not clear which method is truly more correct, as

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

neither will preserve exactly two-body problems at each step. Further, we note that over the span of a picosecond time-step (10^{-12} s), particle dynamics is not truly diffusive but is inertial. Capturing these dynamics would require a different model (e.g. generalized Langevin Dynamics(Van Kampen, 2007)) that tracked both positions and velocities of particles.

Category 2: "Intermediate" cases

This category includes slightly more realistic biological cases with interesting emergent properties. These are particularly useful for illustrating the fundamental conceptual differences among the different modeling approaches.

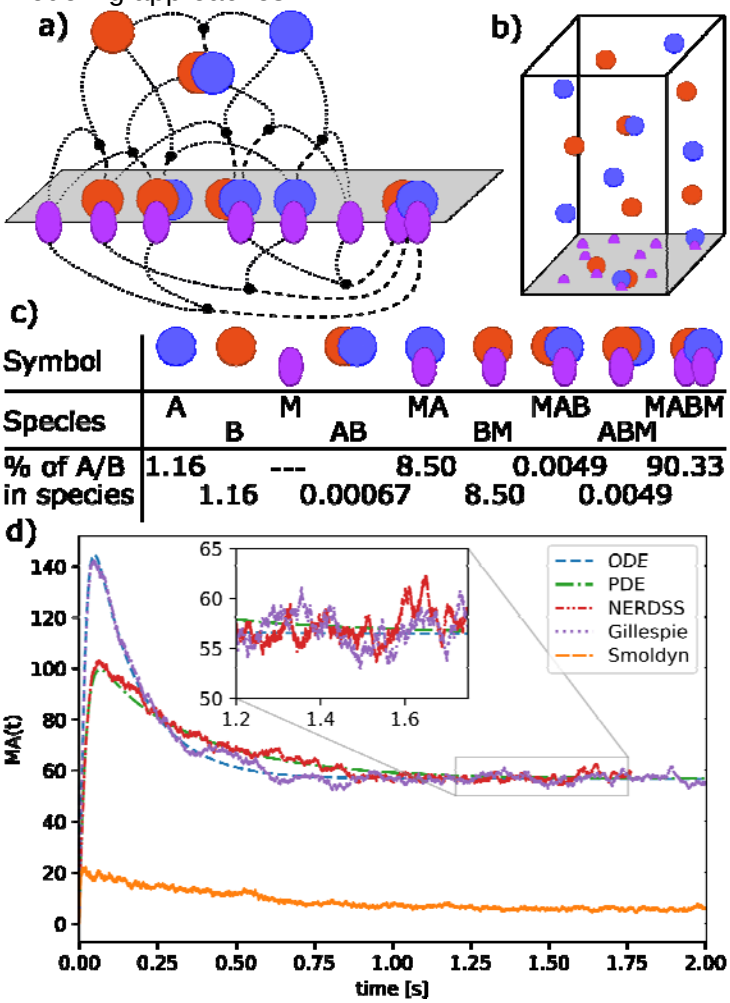


FIGURE 5. Membrane localization can tune speed and stability of protein complex formation. a) A pair of protein binding partners A and B that can also localize to the membrane surface through binding a lipid M can exploit the 2D search space to promote complex formation. There are three binding interactions (below the plane) that involve 2D interactions between protein A and B on the surface, or between a protein and the lipid M. b) The species can partition reversibly between solution and the membrane, forming 9 distinct species, which are listed in (c). c) Model species and fraction of A or B proteins in each at equilibrium. d) Time-dependent formation of a single protein-lipid complex MA. (The time course for BM is identical due to the choice of model parameters.) The initial rise in MA concentration is followed by a drop as the MA and BM complexes combine to form MABM, which dominates at equilibrium. The spatial simulations (PDE and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

NERDSS) exhibit slower MA formation kinetics and a lower peak concentration than the well-mixed simulations (ODE and SSA) due to slower recruitment of proteins to the surface, which is limited by diffusion. Relaxation of MA to equilibrium is also slower for the same reason. For smaller simulation volumes, where diffusion to the surface is fast, these differences disappear. Smoldyn does not reproduce the proper equilibrium due to inaccuracies of binding in 2D.

2A: Exploiting membrane localization to stabilize protein-protein interactions

In a variety of biological processes, including clathrin-mediated endocytosis and initiation of signal transduction, multi-valent proteins localize to membranes and assemble into larger, multi-protein complexes. Reducing dimensionality from a 3D search to a 2D search can accelerate kinetics of receptor binding (Adam and Delbruck, 1968; Berg and Purcell, 1977), and stabilize interactions for macromolecules restricted to the surface (Abel et al., 2012; Kholodenko et al., 2000; Minton, 1995). For cytoplasmic protein binding partners, localization to the 2D membrane can further drive dramatic increases in protein complex stability, as was recently quantified using a simple model of two proteins that bind to one another and to a specific membrane lipid (Yogurtcu and Johnson, 2018) (Fig 5a,b). The origin of the increased stability is largely a concentration effect, where the reactants collide with one another more frequently on the surface than in solution. While a change also must occur in 3D vs 2D equilibrium constants (Wu et al., 2011), the magnitude is usually much smaller than the change in 3D vs 2D search space ($V/A > K^{3D}/K^{2D}$) (Yogurtcu and Johnson, 2018).

With the ability to localize to the membrane and reduce their search space, over 90% of proteins end up bound to one another (Fig. 5c), which contrasts markedly with the purely 3D solution binding result calculated at equilibrium where only 4.6% would end up bound (for K_D=20 μM and $[A]_0=[B]_0$=1 μM). Although all simulation methods are expected to produce the same equilibrium, and in this example stochastic effects are minimal, we find here that spatial effects arise because diffusion slows the localization of A and B molecules to the membrane (Fig. 5d). The height of the box in the spatial simulations is 5 μm, and so despite efficient solution diffusion (50 μm^2/s), the proteins are recruited more slowly to the membrane in spatial simulations than in the well-mixed simulations (compare red and green curves in Fig. 5d with purple and blue ones). The protein-lipid complexes plotted in Fig 5d peak before dropping, as they bind to one another to equilibrate, and the peak is lower when they do not localize rapidly to the membrane. Errors in the results calculated using Smoldyn arise due to approximate treatment of purely 2D interactions, which leads to quantitative deviations from expected equilibria and kinetics for these reactions. This current limitation is being actively addressed in Smoldyn software (S. Andrews, personal communication).

Overall, these types of reactions form a critical component of more complex models of membrane-mediated assembly. Due to the quantitative differences we observe here in time-dependence of species numbers, if the model is coupled to reactions that drive it out of an equilibrium steady-state, this could then drive qualitative changes in the biological outcomes (as in case 2B below).

2B: Increasing stochastic fluctuations in a system with multiple steady states

Positive feedback in combination with other interactions can give rise to systems with multiple, distinct steady states. Such systems can exhibit large differences in their dynamics depending on whether they are simulated deterministically or stochastically. A simple model that illustrates such effects is the

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

autophosphorylating kinase model first introduced by Lisman (Lisman, 1985) and studied more recently by Agarwal et al. (Agarwal et al., 2012), who analyzed a stochastic version of the model. Figure 6a shows a diagram of the model, which consists of a kinase that can activate itself through phosphorylation (reactions 2 and 3) and a phosphatase that can bind and dephosphorylate the active kinase (reactions 4 and 5). Because the production rate of active kinase (Ap) increases with the amount of Ap starting at low concentrations (Fig. 6b, red curve), the system exhibits positive feedback. The rate balance plot shown in Fig. 6b illustrates a case where the production and degradation rates of Ap as a function of Ap intersect multiple times to give rise to multiple steady states. This model has three steady states, also called “fixed points,” two of which are stable (Fig. 6b, filled circles), For a stable fixed point, the balance of production and decay returns the system back to the fixed point following any slight changes in Ap away from the steady state value, whereas at an unstable fixed point the balance of rates carries the system away from the fixed point as a result of any tiny fluctuation.

Deterministic simulations of the system, whether spatial or not, converge rapidly to one of the stable steady states depending on the initial level of Ap (Fig. 6c, black dashed lines). Simulations starting from a state with low initial Ap will reach the steady state with lower Ap and vice versa for high Ap. Stochastic trajectories, on the other hand, may initially stay in the vicinity of the closer steady state, but fluctuations due to noise occasionally induce switches between states. In both trajectories shown in Fig. 6c the system starts in the lower steady state (as indicated by gray shading) but after a minute or so switches to the higher steady state where it continues to display fluctuations, some of which lead to short-lived excursions back to the lower state. For the set of rate parameters we used here, the system spends more total time in the higher state than the lower state (88% vs. 12%). In addition, the length of each segment in the higher state, called the residence time, is longer on average (2.6s vs 0.34s) (see Methods and Fig S1).

Despite the dramatic differences between the deterministic and stochastic simulations for this model, we find the addition of space has relatively modest effects, with spatial models still producing bistable switching as diffusion slows. Fig. 6c shows a non-spatial SSA simulation in the upper panel (Gillespie) and an explicitly spatial simulation in the lower panel (NERDSS). For the geometries and diffusion constant values chosen here, these two simulations yield small differences between the probabilities and residence times for each state (Table S4). When the diffusion constant is reduced by a factor of 10 the probabilities of the lower state drop significantly although not dramatically, and the residence times of both states drop (Fig. S2 and Table S4). The shorter residence times in spatial simulations relative to the non-spatial likely results from larger fluctuations in copy numbers due to transient spatial inhomogeneities, making both steady-states slightly less stable. Reduction of the diffusion coefficient by another factor of 10 makes simulation incompatible with the macroscopic reaction rates specified in the model. In other words, the reaction rate is limited by diffusion (see Box 4). We found that the effect of varying diffusion constant was quantitatively similar in the same model, but with all rates reduced by a factor of 10, reducing the fastest rate from $8*10^8$ to $8*10^7 M^{-1}s^{-1}$ (Table S5). A study on a similar model of bistability found that slow diffusion limited the parameter regimes where one observed bistable switching due to fluctuations in local concentration, but also by changing the effective rates(Endres, 2015). Here we kept the macroscopic rates the same as the non-spatial model even as diffusion constants slow, and our results therefore indicate that timescales are sensitive to spatial inhomogenities that persist longer with slower diffusion.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

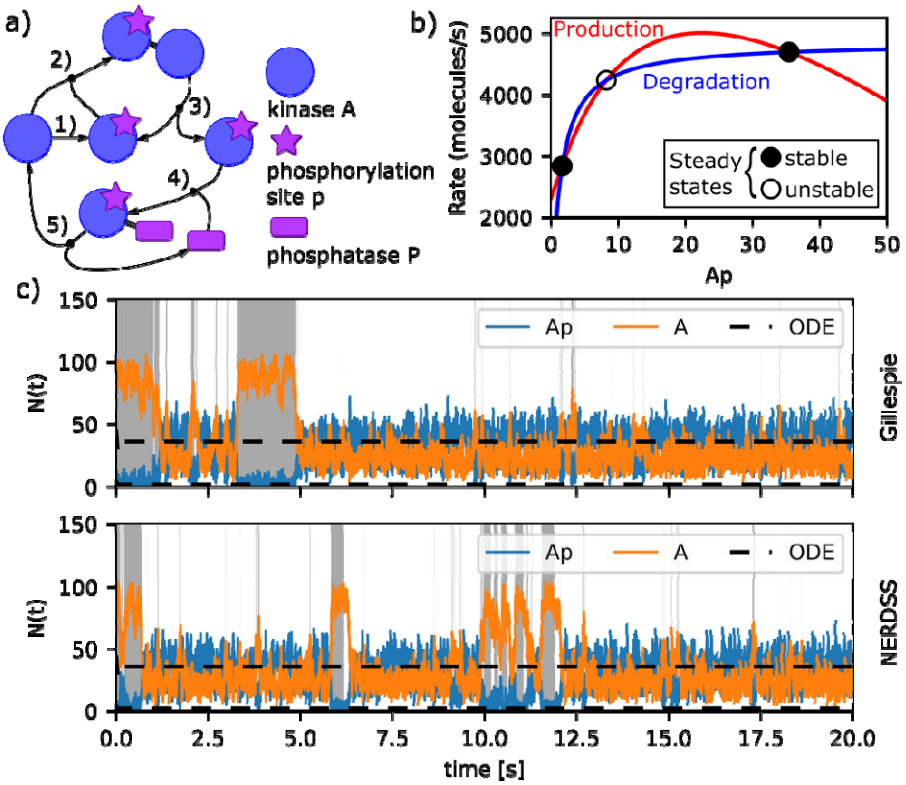


FIGURE 6. Positive feedback induces stochastic switching in a kinase autophosphorylation circuit. (a) Model reactions. The kinase A (blue circle) becomes active upon phosphorylation (purple start; reaction 1) and serves as its own substrate (reactions 2 and 3). A phosphatase P (purple rectangle) binds and dephosphorylates the phosphorylated kinase (reactions 4 and 5). The activation reactions 2 and 3 form a positive feedback for kinase activation. A low rate of spontaneous activation of the kinase (reaction 1) is also included to prevent the system from being trapped in a state with no active kinase. **(b) Rate balance plot identifying steady state concentrations of phosphorylated kinase (Ap).** The red and blue lines show the rates of production (sum of rates of reactions 1 and 3) and degradation (rate of reaction 4) of Ap for the parameter values and initial concentration simulated here (Methods). Intersections of these curves indicate points at which production and degradation rates are equal and hence give rise to a steady state of the system. The two intersection points shown with filled circles indicate the stable steady states of the system, which occur at Ap concentrations of 1.7 and 35.5 molecules respectively. A third steady state (open circle) indicates an unstable steady state, which occurs at a value of 8.3 Ap molecules. **(c) Model trajectories computed with deterministic and stochastic methods.** Deterministic trajectories starting from different initial conditions may relax to either of the two stable steady states (black dashed lines). Stochastic trajectories exhibit fluctuations about each of the steady states and occasionally switch between states with low and high kinase activation as shown by the blue trajectory. The top panel shows the result of a non-spatial stochastic simulation (Gillespie) and the bottom panel shows the result of a spatial stochastic simulation using the same paramenters (NERDSS). Regions of the trajectories shaded gray indicate where the system is in the state with low kinase activation with correspondingly high levels of inactive kinase (orange lines). In the regions with no shading the system is in a state with high kinase activation, as indicated by the active kinase level (blue lines) generally being above the inactive kinase level. The spatial stochastic simulations performed with NERDSS exhibit small differences from the non-spatial simulations carried out using Gillespie when performed in 3D with all species

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

having a diffusion constant of 100 $\mu m^2/s$, and the same macroscopic rate is targeted (see Methods for full details on parameters used in simulation).

Category 3: "Application" cases

Proper implementation of the test cases in this category requires the building blocks described above. These well-defined problems yield rich, biologically-interpretable outputs that are sensitive to parameter choices.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

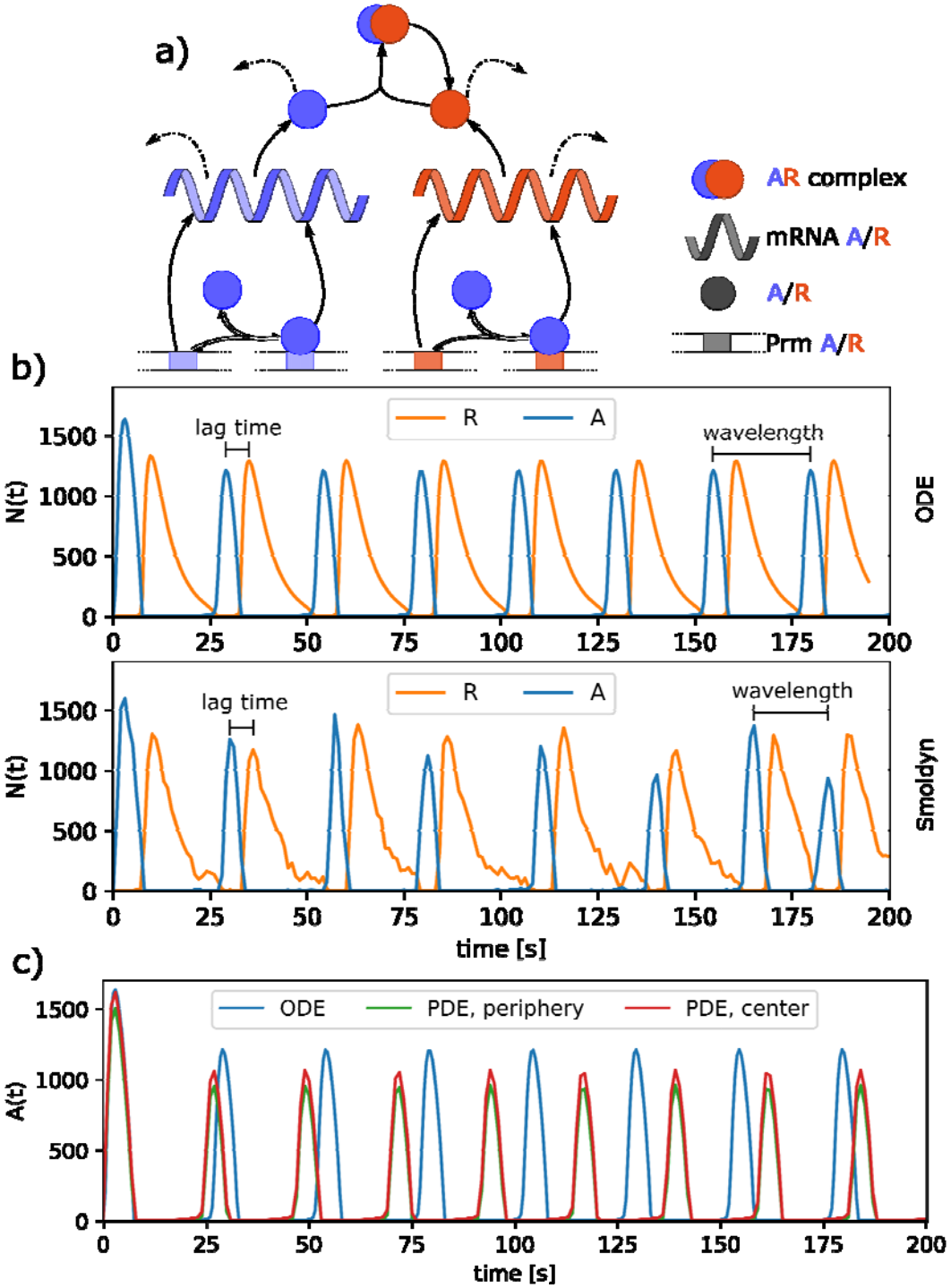


FIGURE 7: Circadian Clock model shows robustness of oscillations to stochastic fluctuations. a) Model of two proteins A and R, translated from corresponding mRNA and transcribed from the corresponding genes (Prm A/R). Protein A is an activator that binds to the promoter of both genes and increases their transcription. Protein R acts as a repressor that binds to A and catalyzes its degradation. All proteins and mRNA also undergo spontaneous degradation with a specific lifetime (dashed arrows). b) Growth of A copy numbers (dashed) peaks and then decreases as it is depleted by repressor R. Initially most R molecules are present in the AR complex, but the number of unbound R molecules increases as A is

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

degraded. Next, the amoung of free R peaks and then begins to decrease because not enough A is present to promote its transcription at a rate that is greater than its rate of spontaneous degradation. The cycle then restarts, with the ODE solution producing highly regular oscillations of 25.2 s between A peaks, and 6s lag between A and R peaks. The stochastic single-particle simulator Smoldyn produces noisier oscillations, but the average behavior is very comparable to the deterministic ODE model, with peak oscillations at 25.9 s and 6.1 s lag between A and R peaks. c) After localization of the promoters to the cell center and slowing of D_A from 10 to 2 $\mu m^2/s$, the A molecules become more concentrated in the cell center (red vs green), which leads to noticeably faster oscillations relative to the non-spatial solution (blue).

3A: Stochastic effects in gene expression

Proteins that are important for controlling circadian clocks exhibit regular oscillations in expression levels. These oscillations are quite robust even in the presence of stochastic copy-number fluctuations. In fact, stochastic fluctuations can support oscillations in regimes that a purely deterministic model cannot (Fig. S3). Using a simple model of circadian oscillations (Vilar et al., 2002) (Fig 7a), we simulated the behavior of an activator protein A and repressor protein R that are produced from mRNA transcribed from a single copy of a gene (one for each protein). Coupling of A and R expression is driven by positive feedback of the activator A, which binds to each gene's promoters to enhance transcription. Protein R also binds to A to effectively degrade it, and all proteins and mRNA are also degraded spontaneously at a constant rate. If the spontaneous degradation rate of protein R is slow, the oscillations will quench in the deterministic model, but persist in the stochastic solutions, as demonstrated in the original work and reproduced in Fig S3.

We find that with the addition of space to the model, with all species diffusing at D=10 $\mu m^2/s$, the oscillation times show no real significant differences in single-particle or deterministic solutions relative to the non-spatial model (Table S6). To quantitatively compare the kinetics across all models, we cannot use simple steady-state values, due to the oscillations. These time-dependent oscillations are nearly perfectly regular in the deterministic models, but are quite imperfect (although recurring) in all stochastic and single-particle methods (Fig. 7b). We therefore measure the average time interval between peaks in the expression of A, and the lag time between the appearance of a peak in A expression followed by a peak in R (see Methods). The similar results across all methods shows how with relatively small spatial dimensions (sphere of R=1 μm), purely 3D reactions, and all species diffusing at the same D=10 $\mu m^2/s$, no spatial dependence was distinguishable.

The lack of any significant spatial dependence is somewhat surprising, because we pushed the reaction rates into the strongly diffusion-influenced regime. The model was actually formulated to describe the slow oscillations of gene expression regulation, with all rates reported in units of hr^{-1}, rendering diffusion times irrelevant. Here, we chose to accelerate the rates by a factor of 3600 (from hr^{-1} to s^{-1}), due to the computational expense of simulating spatial models with explicit diffusion. This result overall indicates that the reactants still mixed faster than any spatial inhomogeneities could emerge, preventing deviations between spatial and non-spatial simulations.

We were finally able to measure a significant dependence on diffusion when we simulated the same set of reactions but we localized and immobilized the promoters for A and R to a small nucleus in the center of the volume, with unrestricted diffusion in and out of this nucleus (Fig 7c). For the same diffusion constants of D=10

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

m^2/s for all other species, this immobilization of promoters was not enough to have an effect, and oscillation times remained not significantly different (Fig S4 and Table S7). However, when the diffusion constant of the activator A was slowed to 2 m^2/s, we observed a persistently higher concentration of A and its mRNA near the cell center relative to the cell periphery (Fig 7 and Fig S4). This localization of A near the gene promoters had the effect of shortening the oscillation period in the PDE from 25s to 22.5s (Table S7). This change was independent of the diffusion constant of R, which does not bind the genes. Further, this same trend was observed in the single-particle simulations, which also produced faster oscillations as D_A slowed to 2 m^2/s (Table S7). Overall, we thus found that while slowing the search time for the Activator for both genes does modulate the oscillation times, the oscillations are quite robust and strongly dictated by total copy numbers and the unimolecular decay reactions, which are inherently independent of diffusion and spatial dimensions. This fairly complicated scenario is one where intuitively we had expected to see strong spatial effects, so it is rather interesting that the actual spatial effects are minor in this case. Certainly this outcome might vary depending on the parameter regimes selected.

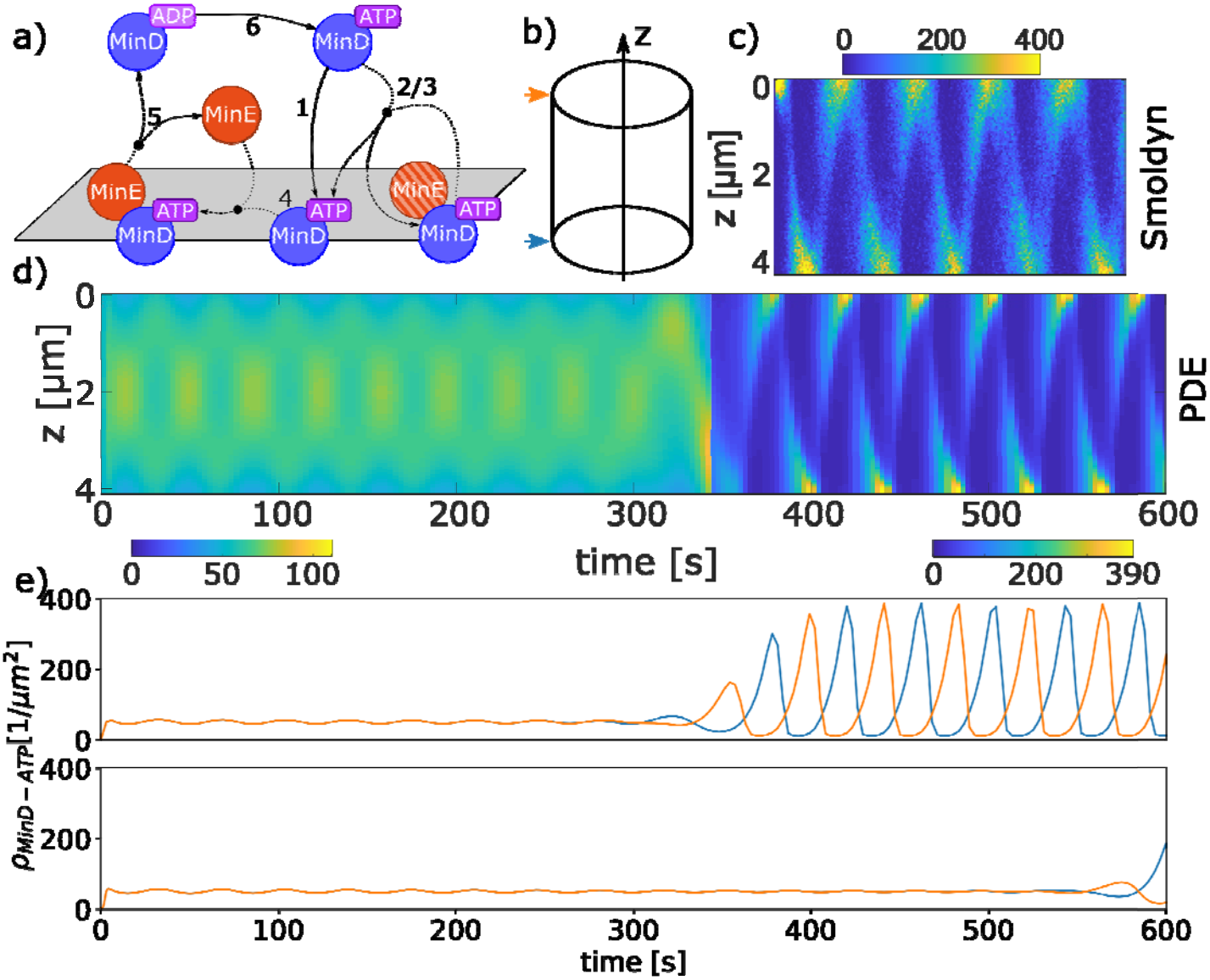


FIGURE 8. Spatial oscillations in protein abundance for the MinCDE cell division model.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

a) The MinD protein can exist in two states, ADP or ATP bound (named MinD-ADP and MinD-ATP). Only MinD-ATP localizes to the membrane in a 3D→2D reaction (producing MinD-ATP^{2D}). Once on the membrane MinD-ATP^{2D} can recruit additional cytoplasmic MinD-ATP (3D→2D) or MinE (3D→2D). The MinD-ATP.MinE^{2D} complex on the membrane recruits MinD-ATP from solution as well (3D→2D), or it dissociates to return MinE to solution and MinD-ATP^{2D} to MinD-ADP. None of these 6 reactions is thus reversible. b) Simulations in a cylindrical cell with L=4um and R=0.5um, with 2.1143μM MinD-ATP and 0.74μM MinE initially well mixed. c-d) The kymographs show how the copy numbers of minDt on the membrane in molecules/µm² oscillate in space and time. c) single-particle simulation Smoldyn d) PDE with uniform initial concentrations has faint symmetric oscillations visible up to ~300s before symmetry breaks. e) Time-dependence of MinD-ATP^{2D} at a point on the left (blue) and right (orange) end of the cell. The oscillations are perfectly symmetric until ~300s in the upper panel. However, if the error tolerance is tightened on the numerical integration, the symmetric oscillations persist longer out to ~550s in the lower panel, illustrating the dependence on numerical precision.

3B: Spatial and temporal oscillations in MinCDE

When coupled to reactions, diffusion that is sufficiently slow or occurs over large enough length scales can establish spatial gradients of concentrations and subsequent oscillations in space and time. By construction, spatial oscillations are lost in non-spatial models, and no temporal oscillations occur either (Fig S5). In a simplified model of bacterial cell division, the MinD and MinE proteins (Huang et al, PNAS 2003) spatially control the location of cell division by creating an oscillating spatial gradient of both proteins in the cytosol and on the membrane (Fig 8). The site of cell division is determined by assembly of a ring constructed from the bacterial tubulin homolog FtsZ. Because the Min proteins inhibit the assembly of the FtsZ ring, the spatial oscillation of the Min proteins from one end to the other in the cylindrical bacterial cell results in the localization of the FtsZ ring to a site very close to the geometrical center (Raskin and de Boer, 1999). This model involves 3D and 3D→2D bimolecular reactions, and unimolecular reactions, with no 2D reactions. Instead of forming explicit polymers, the membrane-bound proteins act to locally increase protein density through recruitment.

We are able to produce very similar spatial and temporal oscillations of the MinDT protein on the membrane in both a stochastic single-particle model (Smoldyn (Andrews, 2017)) and a deterministic PDE solution (using Vcell (Moraru et al., 2008; Schaff et al., 2016)). The major distinction between the two models is that the deterministic PDE is able to support symmetric, or striped oscillations in the protein, as would be generated by two traveling waves in opposite directions. However, we find that these dynamics are not stable even in the deterministic solution—the accumulation of numerical precision error eventually breaks the symmetry and transitions to pole-to-pole oscillations. We show that by increasing the error tolerance on the numerical integration or by increasing the PDE mesh size one can delay the onset of this transition, clearly demonstrating the dependence on numerical precision (Fig 8e and Fig S6). The stochastic single-particle method is therefore not able to support the symmetric oscillations due to copy number fluctuations, and always produces pole-to-pole oscillations (Fig 8b). This shows quite clearly that the pole-to-pole oscillations are more robust to stochastic fluctuations in copy numbers for this geometry, and in the wild-type biological systems, these are the types of oscillations always observed for cells of this length (Raskin and de Boer, 1999).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A challenging aspect of model comparison for any simulations tracking the spatial and time-dependent concentration of species is deciding how best to quantify the results. Analysis is often defined in a problem-specific way. For the MinCDE model, depending on the cell geometry and model parameters, it may not even reach a steady-state over long time scales, but can continue to oscillate and change oscillation amplitude or frequency. For symmetric geometries (like cylinders), we reduced the analysis to a function of time by tracking the MinD-ATP^{2D} molecules at a specific point on the membrane (Fig S7). Then, similar to the analysis for Fig 6, we calculated the average period between oscillations (Methods), finding values of 43.0±0.7s for Smoldyn and 41.3±0.3 s for the PDE.

The oscillations in this model are fairly sensitive to initial concentrations of species (Fig S6) and their relative stoichiometry. If the initial concentrations were cut in half, we found the oscillations disappeared in both models. The MinCDE model has been extensively studied, with the major determining features of oscillations being initial concentrations and cell geometry (see (Halatek and Frey, 2012) and ref therein). The oscillations are relatively robust even to small changes in the reaction network, as long as membrane recruitment (3D→2D) and ATP hydrolysis are included (Halatek and Frey, 2012). We note that for Smoldyn, the oscillations were sensitive to the diffusion coefficients of the membrane bound particles, set here to D=0.05 $\mu m^2/s$. If proteins did not diffuse on the membrane, the oscillations disappeared in Smoldyn, whereas they were unaffected in the PDE. This is in part because the molecules do not have excluded volume in Smoldyn, and because no explicit polymers were formed, they accumulated in a small space rather than spreading along the membrane.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

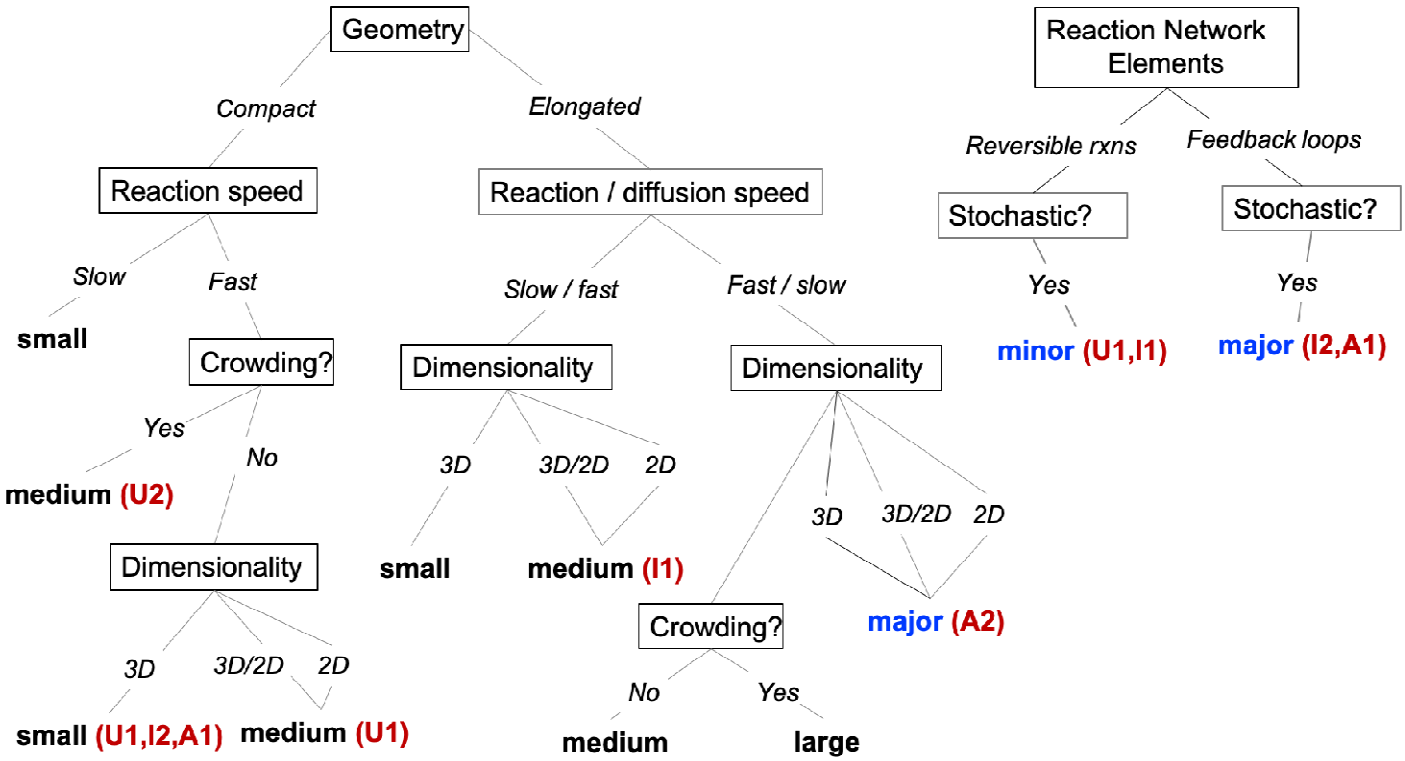


FIGURE 9. Summary of impact of spatial modeling and simulation approaches on quantitative (small, medium, large) and qualitative (major/minor) biochemical behavior. For quantitative behavior, our tree structure separates parameter and geometry regimes to specifically identify the scale of change observed and expected. For qualitative behavior, we note that, especially for reaction-network elements and stochastic effects, observing major changes also depends on additional parameter specifications, but in a less predictable way than the spatial effects (e.g. relative sizes of distinct rates or copy numbers). Minor changes is the default, as we observe for models with purely reversible reactions (rxns). The test cases range from very simple problems (U1 *Bimolecular association in 3D, 2D, and from 3D to 2D*, U2 *Crowding*), via intermediate tests (I1 *Exploiting membrane localization to stabilize protein-protein interactions*, I2 *Increasing stochastic fluctuations in a system with multiple steady states*) to applications that combine different spatial features (A1 *Stochastic effects in gene expression*, A2 *Spatial and temporal oscillations in MinCD*E). Various tools have been applied in the test cases: U1-- ODE, PDE, SSA/Gillespie, particle-based (NERDSS, Smoldyn); U2 – particle-based (NERDSS, eGFRD); I1 – ODE, PDE, SSA /Gillespie, particle-based (NERDSS, Smoldyn); I2 -- ODE, SSA /Gillespie, particle-based (NERDSS); A1 – ODE, PDE, SSA /Gillespie, particle-based (NERDSS, MCell, Smoldyn); A2 – PDE, particle-based (Smoldyn). A detailed description of the theoretical basis and the features offered by these tools/methods can be found in Table S2.

Summary of Test Case Outcomes

By simulating the same biological cases using a range of spatial, non-spatial, stochastic, and deterministic methods, we have shown here how specific biological features and in some cases algorithm selection (e.g. classes of single-particle approaches) can alter quantitative and and even qualitative outcomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

As summarized in Fig 9, quantitative differences can be small, medium, or large based on system geometry and parameter regimes. Qualitative differences can be minor, but can also be major, producing entirely new behavior patterns that emerge. For example, major impacts could be observed in application scenario A2, which combines an elongated geometry with a comparatively slow diffusion (in comparison to the reaction rate), and includes both 3D and 2D dynamics. In general we find that effects of introducing stochasticity to a system may be orthogonal to its spatial properties, as wrongly applying a deterministic method is likely to dominate all other choices. This can be most clearly seen for the results as shown in I2. Overall, we found that, while stochasticity played a minor role in models with reversible interactions (U1, I1), it was capable of driving major changes in models with irreversible interactions and feedback (I2, A2). Quantitatively, crowding might have a large impact, in particular if combined with slow diffusion and high reaction rates. Also the impact on dynamics that involve reactions in 2D, be this 3D→2D or 2D dynamics, is significant and cannot be ignored in many cases. The suitability of specific simulation methods and their limitations have also to be taken into account when any investigator is making choices about which methods to employ for a particular biological problem (see Table S2). For example, if crowding needs to be considered, particle-based approaches, in particular those of the Smoluchowski type, should be applied. However, in the case of very dense crowding, these approaches do not scale because of their computational cost.

The spatial dimension in our test cases primarily drives quantitative changes in outcomes of varying magnitude. However, once multiple features are combined, as in the A2 MinCDE model, major qualitative differences emerge. We expect that our results will extrapolate to informing model selection in a broader range of biological processes. For example, the effect of clustered molecules acts to spatially localize interactions, which similar to the membrane localization (I1) can drive dramatic increases in complex stability, and changes to kinetics. Compartmentalized or highly elongated and narrow geometries also can act to locally alter species concentrations, not unlike the role of dimensionality in driving quantitative changes (U1, I1) or qualitative changes (A2). Lastly, in all the models studied here, all species were initially well-mixed in their system volumes, albeit in U1 (3D→2D) and I1 there was both a membrane and a solution volume. By introducing an explicit spatial concentration gradient via localized sources or sinks of molecules, spatial effects are inevitable, and can only be captured with explicit spatial representations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

DISCUSSION

Since the use of GFP as a genetically encoded fluorescent tag came into widespread use over a quarter-century ago (Chalfie et al., 1994), methods for live-cell light microscopy have advanced spectacularly, and the spatial resolution and temporal resolution now possible for imaging experiments in living cells is truly awe-inspiring (Chen et al., 2014; Guo et al., 2018; Liu et al., 2018). At the same time, a great deal of technical development has gone into the generation of live-cell fluorescent reporters for biological readouts including signaling events, enzymatic activity, and force generation (Greenwald et al., 2018; Yasunaga et al., 2019). Many of the cell biological phenomena illuminated by direct observation of dynamic molecular events in living cells reveal a spatial component that could not be fully appreciated during the era when the most sensitive assays for molecular events were biochemical in nature (Machacek et al., 2009; Tay et al., 2010).

At this point, the quality of time-resolved live cell imaging data for cell biological events, and its quantitative accuracy, is far ahead of our ability to model these events computationally. As a community we could avoid doing the work to develop detailed, spatially resolved models when the data quality was relatively poor, relying on images of fixed cells at poor spatial resolution, but now that the data is excellent there is no excuse. Quantitative data demands quantitative models.

Accurate, spatially-resolved 3D modeling of subcellular processes is inherently challenging, particularly as compared to the larger field of time-resolved cellular modeling where space is not taken into account. In exploring the currently available tools for 3D modeling, and dreaming about what the future might hold, we have identified several sticking points that will require focused community effort to overcome.

Spatially accurate modeling is intrinsically computationally expensive

Most interesting cell biological events unfold over minutes, hours or days, requiring that relevant computational models be able to simulate comparable lengths of real time. For spatially resolved models, keeping track of the positions as well as the numbers and identities of relevant molecular species requires substantially greater computational overhead than simpler models that include only numbers (or concentrations) changing over time. In practice, this difficulty often limits spatially-resolved models to keeping track of only a few molecular species.

An interesting analogy can be drawn to the field of molecular dynamics simulations, which aims to model relative movements of atoms within proteins (Dror et al., 2012), while models of the scales that we are discussing typically aim to simulate the relative movements and transformations of molecules within cells. The equations of motion for atoms within protein molecules are relatively well understood, but accurately calculating their trajectories over even the millisecond time scales relevant for protein folding or conformational changes requires massive computational power (Shaw et al., 2010). Generally, increases in complexity and time scale for simulations of atoms in proteins have benefited from two types of efforts, improvements in hardware including purpose-built computers and distributed computing architectures for parallelization (Snow et al., 2002), accompanied by improvements in statistical sampling approaches to capture rare but significant events (Zuckerman, 2011). Even so, most molecular dynamics simulations still cannot simulate processes central to the lives of proteins such as enzymatic catalysis, because they essentially treat atoms as hard balls subject to defined forces, but ignore the quantum mechanics that would be required to describe chemical transformations. Some hybrid methods have shown promise by including quantum-mechanical detail only in

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the enzyme's active site (Karelina and Kulik, 2017), but naturally this improvement in chemical detail comes at substantial computational cost.

Similarly, for spatially resolved models at the cell biological scale, there is room for substantial improvement in efficient design of hardware and computational parallelization, development of appropriate statistical sampling methods, and hybrid methods that could smoothly integrate full detail in some particular locations or at some particular times with computationally efficient approximations elsewhere (Hellander et al., 2017). GPU architectures offer many advantages for efficient parallelization, and have been effectively used by several groups to push cell biological models to much larger scales (Ghaemi et al., 2020; Hallock et al., 2014), though these approaches have not yet been widely adopted. There are also some kinds of modeling problems that intrinsically cannot be parallelized, where each step in time depends on the state of the system as a whole (Enculescu et al., 2010).

Several features of cell-scale modeling render this a massively more difficult computational challenge as compared to molecular dynamics. One, of course, is the scale; a protein may include thousands or tens of thousands of atoms, while a typical cell will include 2-4 million proteins per cubic micrometer of volume (Milo, 2013), or over 10^{10} in an average mammalian cell, not to mention lipids, nucleic acids, carbohydrates, small metabolites and ions that may all be relevant to propagation of cellular signals. At present, all spatially-resolved models for cell biology therefore necessarily include only a very small sampling of the molecular species that may influence the process. Furthermore, most parameters as fundamental as molecular concentrations and reaction rate constants are not known, or known only approximately (Gutenkunst et al., 2007; Schillings et al., 2015). For that reason, cell biological models must typically also perform a broad sampling of "parameter space", limiting the depth to which any particular instantiation of a spatially-resolved model can be explored.

There is no common framework for sharing and reproducing spatially resolved models

Over the past decade, there has been a humbling realization that a large fraction of published results in experimental biology, including many that have led directly to significant efforts toward drug development, cannot be reproduced (Begley and Ellis, 2012; Prinz et al., 2011). This has led to widespread reform in both the performance and the publication of experimental research, notably with the US National Institutes of Health (NIH) now requiring explicit training in experimental rigor and reproducibility, and altering the structure of research grant applications to address these issues directly (Health, 2018). In order for the modeling enterprise to advance, it will be necessary for our community to establish standards for sharing and reproducing spatially resolved computational models as well.

Sharing and reproducing models is only possible if every quantitative parameter is strictly defined in a meaningful way. This will require a common "language" that is flexible enough to describe a wide variety of complex cell biological problems, and it is particularly challenging to define parameters that can be implemented in an equivalent way for both PDE-based and particle-based simulations of the same process. Even within the confines of our small working group, where we had all agreed to work together closely and were in constant communication throughout the process, we frequently found it challenging to run the "same" model on different platforms because of inconsistent parameter definitions. We also found that published models often lack critical parameters even if they appear at first glance to be completely described.

For time-resolved cell biological models that do not include an explicit spatial component, the Systems Biology Markup Language (SBML) represents a large-scale community effort to improve standardization of

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

model representation. An extension to SBML explicitly designed for spatially-resolved modeling, SBML-spatial, is nearing completion (SBML.org, 2020) and is supported to varying degrees by some modeling systems, although the specification has not yet been formally released and is not widely adopted. Some leaders in the field have begun to organize resources to aid in efforts toward verifying simulation results (reproduciblebiomodels.org, 2020). Several related fields that require highly complex computational models, such as finite element modeling in biomechanics, have also begun to define community standards for reporting and verifying model parameters and results (Erdemir et al., 2012).

Minimal steps toward ensuring reproducibility in cell biological spatial modeling must include a general expectation that all code will be readily accessible (via GitHub or equivalent) and directly runnable by outside users, not dependent on local files or libraries. All code for spatial modeling should include explicitly defined unit tests, such as those we have included here. Furthermore, all parameters used in any model must include meaningful units. This is not a trivial consideration; for example, if the problem being studied involves a transition between a soluble form and a membrane-associated form of a particular molecular species, we have found that rate constants and their units are usually not carefully documented in most published models, and in some cases it is not even clear that that different researchers agree on what the “right” units are for some typical problems.

It is difficult to compare outcomes of different types of models, and to compare model results to experimental data

A typical goal for cell biological modeling is to determine whether a particular underlying mechanistic hypothesis is consistent with experimental observation. For processes with a spatially-dependent stochastic component, which includes many cell biological problems of interest, each computational run (and, indeed, each independent experiment) will give a slightly different result. In order to compare experimental results with simulations, and to compare outcomes of different simulations with one another, the simplest approach is to use some kind of summary statistics that described probabilities of various outcomes. However, it is not often trivial to determine which summary statistics are most appropriate. For example, one influential early model illustrating the importance of stochasticity in genetic regulation leading to cell fate determination was an exploration of the events underlying the switch for bacteriophage lambda from a lysogenic to a lytic state, published in 1998 (Arkin et al., 1998). However, more than a decade passed before a useful method for calculating summary statistics on these extremely rare switching events was developed (Morelli et al., 2009). Ironically, it may in many cases be easier to repeat an experiment for thousands or tens of thousands of cells, for example using automated videomicroscopy (Cai et al., 2018), than to run a computationally intensive spatial simulation the same number of times. In general, the field would benefit from greater attention to scalability in model construction.

Some statistical strategies that were originally developed in other fields have been creatively adapted for applications in spatially-defined stochastic simulations. These include extensions to the singular spectrum analysis methods that are widely used for analysis of time-series data (Shlemov et al., 2015) and the weighted ensemble sampling method that is useful when rare events are of particular interest (Donovan et al., 2016).

Visualization of model outcomes should also be standardized. This is much more conceptually challenging for spatially resolved models than for models where the only output is a change in concentration of molecular species over time. Here recent developments in the field of visual analytics, which combines statistical analysis with visualization and user interactions to cope with large complex data sets, may offer

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

some promise (Matković et al., 2018). Visual analytics takes the human into the loop for exploring data and the space of simulation experiments. Other approaches aim at testing requirements and confirming behavioral expectations automatically. One example of a promising avenue for automatic analysis would be spatio-temporal model checking approaches, which rely on explicitly stating spatio-temporal properties to be checked on the produced trajectories of the spatial simulation (Bartocci et al., 2015).

It may not be possible to address distinct model features required for different kinds of cell biological problems within a single modeling framework

In our comparative work, we emphasized simple biochemical interactions because we were able to implement these examples in comparable ways across multiple different computational platforms. We chose to use only very simple geometries for our example cases, but some of the stochastic simulation platforms we considered are capable of incorporating complex cell geometries, and complex geometries are also standard in most PDE solvers. However, none of the packages we used here are designed to incorporate correct subcellular physics for cytoskeletal mechanics, force generation, fluid flow, or electrostatics. Clearly these physical elements are important in accurate simulations of cellular behaviors. It has been widely recognized for over a century that cellular organization is tightly coupled to cellular structure (Abbot, 1916).

In our opinion, the highest priorities for integration of correct subcellular physics with stochastic molecule-based methods of the kind described here are accurate mechanical models for cytoskeletal filaments, molecular motors, and membranes. Interactions of the cytoskeleton and its associated motors with membranes determine overall cell shape and organelle localization, and these also underlie interactions of cells with one another and with their extracellular matrix. Actions of cytoskeletal filaments and motors directly deform membranes to produce invaginations or protrusions, and membrane surfaces influence cytoskeletal filament growth both because of their mechanical resistance (Keren et al., 2008) and because of their ability to accumulate and spatially organize key regulators (Mullins et al., 2018). Simulation frameworks that could properly deal with these kinds of mechanical and biochemical interactions should be adaptable for a wide variety of related problems, for example exploring the assembly of viral “replication factories” on intracellular membranes of host cells, a process which requires large-scale oligomerization of viral polymerase (Spagnolo et al., 2010).

While we recognize that all models are necessarily approximations, we believe that some approximations are better than others. For membrane mechanics, the Helfrich model (Helfrich, 1973) is the most widely used framework for calculation of energies associated with membrane bending and deformation (Guckenberger and Gekle, 2017). While the theory itself is simple and elegant, calculation of energies for highly complex geometries in a continuum mechanics model can become complicated. A few interesting approaches have been used to enable cellular geometries to adapt in response to biochemical reactions (Angermann et al., 2012; Tanaka et al., 2015), but so far none of them have attempted to incorporate correct Helfrich-derived bending energies. PDE-based models for membrane-protein interactions can better approximate the correct underlying physics, but have so far been limited to fairly simple geometries (Rangamani et al., 2014; Wu et al., 2018). Ultimately, the membrane mechanics must be coupled to the proteins and protein assemblies that provide most of the force necessary to induce membrane bending and remodeling.

Capturing the dynamics of protein assemblies, both as highly ordered filaments or lattices (e.g. microtubules or clathrin-coated vesicles) and as disordered condensates, will be essential for bridging

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

biochemical interactions with mechanics. Recent advances in extending single-particle methods to include rigid (Varga et al., 2020) or polymer-like structures (Hoffmann et al., 2019; Michalski and Loew, 2016) have enabled simulations of highly ordered clathrin lattices and viral shells (Varga et al., 2020), or disordered condensate-like assemblies (Chattaraj et al., 2019), respectively. For cytoskeletal mechanics, in many cases a simple approximation of cytoskeletal filaments as elastic rods and motor proteins as discrete localized generators of quantized directional force can give impressively realistic simulation outcomes (Nedelec and Foethke, 2007b). This approach has been widely used for studies of dynamic microtubule and actin structures in eukaryotic cells (Karsenti et al., 2006; Odell and Foe, 2008; Rubinstein et al., 2009). It has recently been demonstrated that mechanically accurate modeling of cytoskeletal filament dynamics can be fruitfully combined with a continuum model for membrane bending in order to simulate endocytosis (Akamatsu et al., 2020).

More generally, it is clear that much more work needs to be done in the development of multi-scale simulations and hybrid computational approaches (Groen et al., 2019). For example, in some intermediate regimes approximate subvolume-based simulations may be able to bridge the gap between well-mixed and particle-based simulations. It will require careful work to understand what approximations are generally useful when jumping across scales in hybrid models, and what approximations give rise to propagating errors.

Overall the members of our working group have thoroughly enjoyed this opportunity to critically evaluate the state of tools that are currently available for simulation of cell biological processes. Although we have of course identified important gaps, we are generally optimistic about recent developments and look forward to incorporation of more sophisticated computational methods to improve the efficiency and accuracy of cell biological modeling as these methods are developed. Spatially accurate cellular simulation is a field still in its infancy, but its future is bright.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

METHODS

Simulation Solvers:

ODE, PDE, and Gillespie simulations were all run using Virtual Cell, version 7(Moraru et al., 2008). Unless otherwise noted, the PDE solver used a fully-implicit finite volume, regular grid (variable time step), max step of 0.1 s. Absolute error tol: 10^{-9}, relative error: 10^{-7}. The mesh used the default values, which typically resulted in mesh side lengths of ~0.02 μm. We note that for the PDE, the mesh, volume, or concentrations sometimes had to be adjusted slightly, to ensure the copy numbers were the same for accurate comparisons. The ODE used the Combined Stiff Solver (IDA/CVODE), abs error tol: 10^{-9}, relative error tol: 10^{-9}. Gillespie simulations used the Gibson-Bruck method, an exact stochastic simulation algorithm, implemented in Virtual cell, or for the auto-phosphorylation model, RuleBender(Smith et al., 2012).

All single-particle solvers propagated Brownian dynamics, with no additional forces added. Smoldyn simulations were run either through Virtual Cell, version 7, or using the stand-alone Smoldyn software, version 2.60/2.61(Andrews, 2017). FPR or NERDSS simulations were run using the NERDSS software, version 1(Varga et al., 2020), which uses the FPR algorithms(Johnson, 2018; Johnson and Hummer, 2014; Yogurtcu and Johnson, 2015) to solve the single-particle reaction-diffusion model. GFRD simulations were run using eGFRD software(Sokolowski et al., 2019) from August 3rd 2019. MCell simulations were run using MCell Version 3.3(Kerr et al., 2008).

Models and simulation parameters:

Unless otherwise noted, species in spatial simulations were initialized as uniformly distributed.

1. Units Tests:

1a. 3D Reversible binding A+B ⇌ C. Model parameters: V=3.1934 μm^3 cube, $[A]_0=[B]_0$=52 μM, or 100,000 copies, $[C]_0$=0. k_{on}=1.476 x 10^7 $M^{-1}s^{-1}$, k_{off}=0.02451 s^{-1}. Microscopic rates, k_a=1000 $nm^3\mu s^{-1}$, k_b=1s^{-1}. K_D=0.00166 μM. For FPR/NERDSS, σ=1 nm. $D_A=D_B=D_C$=1 μm^2/s. $[A]_{eq}$= 0.293 μM, or 563.5 copies, reached after ~1 s. *Sim parameters:* Single-particle methods $\Delta t=10^{-7}$-10^{-6} s. FPR/NERDSS N_{traj}=18, Gillespie N_{traj}=10, and Smoldyn N_{traj}=1.

1b. 2D Reversible binding A+B ⇌ C. Model parameters: A=1 μm^2, flat surface. $[A]_0=[B]_0$=1000 μm^{-2}, or 1000 copies, $[C]_0$=0. k_{on}=3.07 $\mu m^2 s^{-1}$, k_{off}=3.07 s^{-1}. Microscopic rates, k_a=10 $\mu m^2 s^{-1}$, k_b=10 s^{-1}. K_D=0.00166 μM. For FPR/NERDSS, σ=1 nm. $D_A=D_B$ =1 μm^2/s, D_C=0.5 μm^2/s. $[A]_{eq}$= 31.1267 μm^{-2}, reached after ~0.1 s. *Sim parameters:* Single-particle methods $\Delta t=10^{-7}$s. FPR/NERDSS N_{traj}=5, Gillespie N_{traj}=5, Smoldyn N_{traj}=2.

1c. 3D→2D Reversible binding A+B ⇌ C. Model parameters: V=1 μm^3 cube, flat membrane surface of 2.2 μm x 2.2 μm. $[A]_0$=1 μM, or 602 copies, $[B]_0$=6045 μm^{-2}, or 29,258 copies, $[C]_0$=0. k_{on}=8.44 x 10^7 $M^{-1}s^{-1}$, k_{off}=70.085 s^{-1}. Microscopic rates, k_a^{3D}=500 $nm^3\mu s^{-1}$, k_b=250 s^{-1}. K_D=0.83 μM. σ=1 nm. D_A =30 μm^2/s, D_B=1 μm^2/s, D_C=0.97 μm^2/s. $[A]_{eq}$= 10.3 copies, or 0.017 μM, reached after ~0.02 s. *Sim parameters:* Single-particle methods $\Delta t=10^{-7}$s. FPR/NERDSS N_{traj}=10, Gillespie N_{traj}=5, and Smoldyn N_{traj}=3. Smoldyn $\Delta t=10^{-6}$ s, with the longer steps converging to the PDE solution.

1d. Crowding A+B→B+C. Model Parameters: V=12,500 nm^3 Cube (side length of 23.21nm), periodic boundary conditions enforced. [B]=100 copies (13.3 mM). $[A]_0$=100 copies. $[C]_0$ varied across 6 systems: [0,

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

994, 2187, 3381, 4575, 5768], corresponding to total volume fractions occupied of [0.00838, 0.05, 0.1, 0.15, 0.2, 0.25]. Reactions: 1. A+B→B+C k_a=85 nm^3/μs, 2. A+A k_a=0, 3. A+C k_a=0, 4. B+B k_a=0, 5. B+C k_a=0, 6. C+C k_a=0. With *no volume exclusion,* Rxn 1 has k_{on}=63.5 nm^3/μs = 3.82 x 10^7 M^{-1}s^{-1}. σ=1 nm, and thus each particle is effectively modeled as a volume-excluding sphere of diameter=1 nm. D_A = D_A = D_c = 10 μm^2/s. *Sim parameters:* For FPR/NERDSS(Varga et al., 2020), a maximal Δt for each system was limited by the density, values for each crowding fraction: [10^{-11} s, 10^{-11} s , 10^{-11} s, 5x10^{-12} s, 5x10^{-12} s, 10^{-12} s]. N_{traj} =40-80 for each crowding fraction. Another N_{traj} =40-80 were run at another time-step (5 times faster) to verify quantitatively similar results. For eGFRD(Sokolowski et al., 2019), N_{traj} =10 for each crowding fraction. Algorithmic adjustments to standard FPR: First, when an A+B→B+C reaction occurred, the products were kept at their same coordinates as the reactants. Second, at each time-step, particle positions are updated, and they must avoid overlapping (r≥σ) all other particles, considering all particles within collision distance (r< R_{max})(Johnson and Hummer, 2014). Due to the high crowding, we used a more stringent criteria to update particle positions. We defined clusters that contained all the particles with a capacity to possibly overlap all their partners. The positions of all particles in this cluster were simultaneously updated to avoid any overlap. For large clusters, to reduce the number of rejected updates, the time-step was reduced by a factor of 10 and then updated 10 times, to the full Δt.

2. Intermediate Tests:

2a Membrane Localization Model. Model Parameters: V=1.1045 μm^3, A=0.2209 μm^2, rectangular Cube, 0.47 μm x 0.47 μm surface, height=5μm. $[A]_0$=$[B]_0$=1 μM or 665 copies, $[M]_0$=5.6474μM=17000 /μm^2, or 3755 copies. $[AM]_0$=$[BM]_0$=$[AB]_0$=$[ABM]_0$=. $[MAB]_0$=$[MABM]_0$=0. Reactions: 1. A+B⇌AB k_{on}=0.05 μM^{-1}s^{-1}. 2. A+M⇌AM k_{on}=2 μM^{-1}s^{-1}. 3. B+M⇌ BM k_{on}=2 μM^{-1}s^{-1}. 4. A+BM ⇌ABM k_{on}=0.05 μM^{-1}s^{-1}. 5. B+MA⇌MAB k_{on}=0.05 μM^{-1}s^{-1}. 6. AB+M ⇌ABM k_{on}=2 μM^{-1}s^{-1}. 7. AB+M⇌MAB k_{on}=2 μM^{-1}s^{-1}. 8. MA+BM⇌MABM 0.0415μm^{2}s^{-1}. 9. MAB+M⇌MABM 1.6693μm^{2}s^{-1}. 10. M+ABM ⇌MABM 1.6693μm^{2}s^{-1}. For all reactions, k_{off}=1 s^{-1}. At equilibrium, 90.3% of proteins were in complex with one another on the membrane (MABM), or 600 copies, and 1.16% were remaining in solution, or 7.7 copies. σ=1 nm for all rxns, D_A = D_B =50 μm^2/s. D_{AB} = 25μm^2/s. D_{AM} = D_{BM}=0.495μm^2/s. D_{ABM} = D_{MAB}=0.249μm^2/s. D_{MABM}=0.248μm^2/s D_M=0.5 μm^2/s. All membrane bound had D_z=0.
Sim parameters: Single-particle methods Δt=10^{-7}s. FPR/NERDSS N_{traj}=10. Gillespie N_{traj} =10 Smoldyn N_{traj} =10. We note that the model includes 10 coupled pairwise reactions (three are 3D, four are 3D→2D, and three are 2D), although only 6 equilibrium constants are unique. This is because there are only 3 reactants and multiple thermodynamic cycles that can result in the final assemblies (see e.g. MABM in Fig 5a). To preserve a thermodynamic equilibrium at steady-state, not all rates can be defined independently.

2b Auto Phosphorylation: Model Parameters A: V=0.003 μm^3 . $[A]_0$=103 copies, $[Ap]_0$=5 copies, $[P]_0$=9 copies. $[A_Ap]_0$=0, $[Ap_P]_0$=0. 1. A→Ap at 21.2 s^{-1}. 2. A+Ap→A_Ap at 10^{7}M^{-1}s^{-1}, k_a=16.7 nm^3μs^{-1}, σ=1 nm. 3.A_Ap→Ap+Ap at 200 s^{-1}. 4. Ap+P→Ap_P at 8*10^{8}M^{-1}s^{-1}, k_a=2820 nm^3μs^{-1}, σ=1 nm and 5. Ap_P→A+P at 539 s^{-1}. D_A = D_P =100 μm^2/s. A second spatial stochastic simulation had D_A = D_P =10 μm^2/s, which required new bimolecular parameters for rxn 2: k_a=17.8 nm^3μs^{-1}, σ=1 nm, rxn 4: k_a=11191 nm^3μs^{-1}, σ=6 nm. NOTE: the binding radius had to be increased to recover the large macroscopic binding rates for rxn 4. A third spatial stochastic simulation had D_A = D_P =40 μm^2/s, which required new bimolecular parameters for rxn 2: k_a=17.2 nm^3μs^{-1}, σ=1 nm, rxn 4: k_a=3919 nm^3μs^{-1}, σ=4 nm.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Sim Parameters: Single particle methods: $\Delta t = 2*10^{-8}$s, N_{traj} =2. For D=10, $\Delta t = 1*10^{-7}$s, N_{traj} =2. For D=40, $\Delta t = 1*10^{-7}$s, N_{traj} =2. All simulations were >=200s to obtain statistics on switching times.

B version of model with all rates reduced by 10: 1. 2.12 s^{-1}. 2. $10^6 M^{-1}s^{-1}$, k_a=1.67 $nm^3 \mu s^{-1}$, σ=1 nm. 3. 20 s^{-1}. 4. $8*10^7 M^{-1}s^{-1}$, k_a=282.0 $nm^3 \mu s^{-1}$, σ=1 nm and 5. 53.9 s^{-1}. $D_A = D_P$ =10 $\mu m^2/s$. A second spatial stochastic simulation had $D_A = D_P$ =100 $\mu m^2/s$, which required for rxn 4: k_a=140.31 $nm^3 \mu s^{-1}$, σ=1 nm. A third spatial stochastic simulation had $D_A = D_P$ =1 $\mu m^2/s$, which required for rxn 2: k_a=1.78 $nm^3 \mu s^{-1}$, σ=1 nm. rxn 4: k_a=1119.14 $nm^3 \mu s^{-1}$, σ=6 nm.
Sim Parameters: $\Delta t = 2*10^{-7}$s, N_{traj} =2. For D=100, $\Delta t = 2*10^{-8}$s, N_{traj} =2. For D=1, $\Delta t = 1*10^{-6}$s, N_{traj} =2. All statistics were collected for at least 2000s, except the fast diffusion constant. With only 100s for D=100, we observed transitions but insufficient statistics to determine trends relative to the other models.

3. Application Tests:
3a Circadian Clock Oscillator Model. Model Parameters A: 4.189 μm^3, cube, or a sphere with R=1 μm. $[prmA]_0=[prmR]_0$=1 copy, or 0.000397 μM. $[A]_0=[R]_0=[C]_0$= $[mRNAA]_0$= $[mRNAR]_0$= $[prmA.A]_0=[prmR.A]_0$=0. Reactions: 1. A + R → C k_{on}=1204 $uM^{-1}s^{-1}$, k_a=356,263 $nm^3 \mu s^{-1}$, σ=8 nm 2. prmA → prmA + mRNAA k_2=50 s^{-1}. 3. prmA.A → prmA.A + mRNAA k_3=500 s^{-1}. 4. prmR → prmR + mRNAR k_4=0.01 s^{-1}. 5. prmR.A → prmR.A + mRNAR k_5=50 s^{-1}. 6. prmA + A ↔ prmA.A k_{6f}=602 $uM^{-1}s^{-1}$, k_{6b}=50 s^{-1}, k_a=4889 $nm^3 \mu s^{-1}$, k_b=244.5 s^{-1} σ=5 nm 7. prmR + A↔ prmR.A k_{7f}=602 $uM^{-1}s^{-1}$, k_{7b}=100 s^{-1}, k_a=4889 $nm^3 \mu s^{-1}$, k_b=489 s^{-1} σ=5 nm 8. mRNAA → mRNAA + A k_8=50 s^{-1}. 9. mRNA_R → mRNA_R + R k_9=5 s^{-1}. 10. mRNA_A → NULL k_{10}=10 s^{-1}. 11. mRNA_R → NULL k_{11}=0.5 s^{-1}. 12. A → NULL k_{12}=1 s^{-1}. 13. R→ NULL k_{13}=0.2 s^{-1}. 14. C→ R k_{14}=1 s^{-1}. D=10 $\mu m^2/s$ (for all 9 species).
B version of the model was run with one modification from the original A model: 13. R→ NULL k_{13}=0.05 s^{-1}, eliminating oscillations in the deterministic model Fig S3.
C version of this model was created with modifications from original A model: a) promoters were spatially localized to a sphere at the center of the volume of R=0.1 μm. b) $D_{prmA}=D_{prmR}=D_{prmA.A}=D_{prmA.R}$=0. c) $[prmA]_0=[prmR]_0$=1 copy, or here 0.3964 μM. No barrier existed for proteins to access this 'nucleus' in the cell center. Full details in SI Methods.

Sim parameters: NERDSS Δt=0.5, 2, and 10 μs, and Smoldyn Δt=100 μs . Gillespie N_{traj}=10. Smoldyn N_{traj}=2, NERDSS N_{traj}=2 (600s each) see Table S6 and S7. Time-steps were limited by the fast unbinding reaction of the single promoter, Rxn 7.

3b Model: MinCDE Cell division Model. Model Parameters: Cylindrical volume, h=4 μm, R=0.5 μm. Five species: $[MinDT]_0$= 2.1143 μM, or 4000 copies. $[MinE]_0$=0.74 μM, or 1400 copies. $[MinDD]_0=[minDt]_0=[minDt.minE]_0$=0, where lower case indicates membrane bound. Six irreversible reactions. 1. MinDT→minDt, k_1=0.025 um.s^{-1}. 2. MinDT + minDt→2minDt, k_2=0.903s^{-1}.uM^{-1}. 3. MinDT+minDt.minE→minDt + minDt.minE, k_3=0.903 s^{-1}.uM^{-1}. 4. MinE+minDt→minDt.minE, k_4=56.0 s^{-1}.uM^{-1}. 5. minDt.minE→MinDD + MinE, k_5=0.7s^{-1}. 6. MinDD→MinDT, k_6=1s^{-1}. *D*=2.5 $\mu m^2/s$ for MinDT, MinDD, and MinE. D=0 $\mu m^2/s$ for minDt and minDt.minE in the PDE. Necessary modification for Smoldyn to produce oscillations: *D*=0.05 $\mu m^2/s$ for minDt and minDt.minE, initial copies 2000 MinDT, 2000 minDt.
Sim parameters: Smoldyn: N_{traj} =4, Δt =2ms.

Analysis Methods:

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Equilibria and error: For all stochastic methods, we calculated averages and variances of copy numbers in time, over *N* trajectories. We used the standard error of the mean at time points: SEM(*t*)=$\langle(A(t)-\langle A(t)\rangle_N)^2\rangle_N/\sqrt{N}$. For equilibrium values, we used differences from the mean to determine when equilibrium was reached. We note that for all models, we selected parameters such that the equilibrium was not too close to zero, which is important for assessing clear deviations from the proper equilibrium values.

Clock Oscillator timescales: The average time separation between A peaks, between R peaks, and the lag time between A and R peaks was calculated in two ways. First, by finding the peak maxima and simply calculating the separation between neighboring peaks, this was then averaged over all peaks. Second, to calculate time period of peaks, we used a discrete FFT, that was zero-padded out to 5000 s to increase the resolution of the frequencies sampled. The frequency with the largest coefficient, f_{max}, was used to define the period of the oscillation ($1/f_{max}$). The cross-correlation between the A and R time-series was used to identify the lag-time, again keeping the time with the largest correlation coefficient. Both methods produced similar results, and were performed using MATLAB functions. We report in SI tables the periods determined from the find peaks method, with SEMs reporting the calculated $\sigma/\sqrt{N_{peak}}$.

Fit to kinetics of crowding simulations: To define the k_{fit} values for the crowding simulations as a function of packing fraction, we fit the copies of A vs time $A(t)=A_0\exp(-k_{fit}B_{tot}t)$, where B_{tot} is 0.008 /nm^3, and we performed nonlinear fitting to the exponential, rather than the logarithm of A(t). We fit each individual trajectory, and report the mean of these values $\langle k_{fit}\rangle=\sum_{n=1}^{N}k_{fit}(n)$, where N is total trajectories. The SEM is calculated based on the variance around this mean value. Fits to the averaged trajectory gave similar values, and fits using the ln(*A*(*t*)) produced the same trends, but with slightly lower k_{fit} values. The same approach was used for both FPR and GFRD trajectories.

MinCDE timescales: For the MinCDE models, we calculated the time-scales of oscillations by choosing fixed points in space at either end of the cylinder (Fig S7). Once the oscillations were pole-to-pole, we quantified the period of oscillations in time using the same approach as applied to the clock oscillator model above, by finding the peak maxima and taking the average of time separations between adjacent peaks. This produced periods of 43.0±0.7 s for Smoldyn, and 41.3±0.3 s for the PDE.

Autophosphorylation State Assignments: To assign stochastic trajectories to states, 1=low phosphorylation, 2=high phosphorylation, we report results based on thresholding the values of A(t) and Ap(t). We created a transition region, where points could be in either 1 or 2 depending on their previous state, to minimize rapid re-crossings between states (see Fig S1). If [A(t)>62+Ap(t)], assign to state 1, elseif [A(t) < 35+Ap(t)] assign state 2, else [if(State(t-1)=*i*, State(t)=*i*], where *i*=[1,2]. We then counted the number of times in each state, length of intervals spent in each state, and transitions between points separated by a time δt to construct $p_{trans}(\delta t)=\begin{bmatrix}p_{11} & p_{12}\\ p_{21} & p_{22}\end{bmatrix}$. We calculated $p_{state,i}=n_{state,i}/(n_{state,1}+n_{state,2})$, and $\bar{\tau_i}=\frac{1}{N}\sum_{n=1}^{N}\tau_i(n)$, where *N* is the number of intervals spent in state i, each of duration τ_i. Distributions of these residence times are shown in Fig S1. To calculate errors on the probability of each state, each trajectory was split into 10 chunks (length at least the larger residence time), and these 10 state probabilities were used to calculate a mean and standard deviation/SEM.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ACKNOWLEDGMENTS: This work was funded by the National Institute for Mathematical and Biological Synthesis (NIMBioS). MEJ also gratefully acknowledges computing resources from JHU's MARCC supercomputer and XSEDE through XRAC MCB150059, and support from an NIH MIRA R35GM133644. **M.M.S. and T.P. acknowledge support by the intramural research program of the National Institute of Allergy and Infectious Diseases, NIAID, NIH.**

REFERENCES

1. Abbot, E.S. (1916). The Causal Relations between Structure and Function in Biology. The American Journal of Psychology *27*, 245-250.
2. Abel, S.M., Roose, J.P., Groves, J.T., Weiss, A., and Chakraborty, A.K. (2012). The membrane environment can promote or suppress bistability in cell signaling networks. J Phys Chem B *116*, 3630-3640.
3. Adam, G., and Delbruck, M. (1968). Reduction of Dimensionality in Biological Diffusion Processes. In Structural Chemistry and Molecular Biology (San Francisco: Freeman), pp. 198-215.
4. Agarwal, A., Adams, R., Castellani, G.C., and Shouval, H.Z. (2012). On the precision of quasi steady state assumptions in stochastic dynamics. J Chem Phys *137*, 044105.
5. Akamatsu, M., Vasan, R., Serwas, D., Ferrin, M.A., Rangamani, P., and Drubin, D.G. (2020). Principles of self-organization and load adaptation by the actin cytoskeleton during clathrin-mediated endocytosis. Elife *9*.
6. Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., and Sorger, P.K. (2006). Physicochemical modelling of cell signalling pathways. Nat Cell Biol *8*, 1195-1203.
7. Amonlirdviman, K., Khare, N.A., Tree, D.R., Chen, W.S., Axelrod, J.D., and Tomlin, C.J. (2005). Mathematical modeling of planar cell polarity to understand domineering nonautonomy. Science *307*, 423-426.
8. Ando, T., and Skolnick, J. (2010). Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proceedings of the National Academy of Sciences of the United States of America *107*, 18457-18462.
9. Andrews, S.S. (2017). Smoldyn: particle-based simulation with rule-based modeling, improved molecular interaction and a library interface. Bioinformatics *33*, 710-717.
10. Andrews, S.S. (2020). Effects of surfaces and macromolecular crowding on bimolecular reaction rates. Phys Biol *17*, 045001.
11. Andrews, S.S., Addy, N.J., Brent, R., and Arkin, A.P. (2010). Detailed Simulations of Cell Biology with Smoldyn 2.1. PLOS Computational Biology *6*, e1000705.
12. Andrews, S.S., and Bray, D. (2004). Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Physical Biology *1*, 137–151.
13. Angermann, B.R., Klauschen, F., Garcia, A.D., Prustel, T., Zhang, F., Germain, R.N., and Meier-Schellersheim, M. (2012). Computational modeling of cellular signaling processes embedded into dynamic spatial contexts. Nature Methods *9*, 283-289.
14. Arkin, A., Ross, J., and McAdams, H.H. (1998). Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. Genetics *149*, 1633-1648.
15. Bartocci, E., Bortolussi, L., Milios, D., Nenzi, L., and Sanguinetti, G. (2015). Studying Emergent Behaviours in Morphogenesis Using Signal Spatio-Temporal Logic (Springer International Publishing).
16. Bartocci, E., and Lio, P. (2016). Computational Modeling, Formal Analysis, and Tools for Systems Biology. PLoS Comput Biol *12*, e1004591.
17. Begley, C.G., and Ellis, L.M. (2012). Raise standards for preclinical cancer research. Nature *483*, 531-533.
18. Berg, H.C., and Purcell, E.M. (1977). Physics of Chemoreception. Biophysical Journal *20*, 193-219.
19. Bortz, A.B., Kalos, M.H., and Lebowitz, J.L. (1975). A new algorithm for Monte Carlo simulation of Ising spin systems. Journal of Computational Physics *17*, 10-18.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

20. Cai, Y., Hossain, M.J., Hériché, J.-K., Politi, A.Z., Walther, N., Koch, B., Wachsmuth, M., Nijmeijer, B., Kueblbeck, M., Martinic-Kavur, M.*, et al.* (2018). Experimental and computational framework for a dynamic protein atlas of human cell division. Nature *561*, 411-415.
21. Cao, Y., Terebus, A., and Liang, J. (2016). State Space Truncation with Quantified Errors for Accurate Solutions to Discrete Chemical Master Equation. Bulletin of Mathematical Biology *78*, 617-661.
22. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.
23. Chattaraj, A., Youngstrom, M., and Loew, L.M. (2019). The Interplay of Structural and Cellular Biophysics Controls Clustering of Multivalent Molecules. Biophys J *116*, 560-572.
24. Chen, B.-C., Legant, W.R., Wang, K., Shao, L., Milkie, D.E., Davidson, M.W., Janetopoulos, C., Wu, X.S., Hammer, J.A., Liu, Z.*, et al.* (2014). Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution. Science *346*.
25. Chen, Y.E., Tropini, C., Jonas, K., Tsokos, C.G., Huang, K.C., and Laub, M.T. (2011). Spatial gradient of protein phosphorylation underlies replicative asymmetry in a bacterium. Proc Natl Acad Sci U S A *108*, 1052-1057.
26. Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412-415.
27. Cohen, J.E. (2004). Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. PLoS Biol *2*, e439.
28. Collins, F.C., and Kimball, G.E. (1949). Diffusion-Controlled Reaction Rates. Journal of Colloid Science *4*, 425-437.
29. Cugno, A., Bartol, T.M., Sejnowski, T.J., Iyengar, R., and Rangamani, P. (2019). Geometric principles of second messenger dynamics in dendritic spines. Scientific Reports *9*, 11676.
30. Dawes, A.T., and Munro, E.M. (2011). PAR-3 oligomerization may provide an actin-independent mechanism to maintain distinct par protein domains in the early Caenorhabditis elegans embryo. Biophys J *101*, 1412-1422.
31. Donovan, R.M., Tapia, J.-J., Sullivan, D.P., Faeder, J.R., Murphy, R.F., Dittrich, M., and Zuckerman, D.M. (2016). Unbiased Rare Event Sampling in Spatial Stochastic Systems Biology Models Using a Weighted Ensemble of Trajectories. Plos Comput Biol *12*, e1004611.
32. Dror, R.O., Dirks, R.M., Grossman, J.P., Xu, H., and Shaw, D.E. (2012). Biomolecular Simulation: A Computational Microscope for Molecular Biology. Annual Review of Biophysics *41*, 429-452.
33. Enculescu, M., Sabouri-Ghomi, M., Danuser, G., and Falcke, M. (2010). Modeling of Protrusion Phenotypes Driven by the Actin-Membrane Interaction. Biophysical Journal *98*, 1571-1581.
34. Endres, R.G. (2015). Bistability: requirements on cell-volume, protein diffusion, and thermodynamics. PLoS One *10*, e0121681.
35. Erban, R., and Chapman, S.J. (2009). Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. Phys Biol *6*, 046001.
36. Erdemir, A., Guess, T.M., Halloran, J., Tadepalli, S.C., and Morrison, T.M. (2012). Considerations for reporting finite element analysis studies in biomechanics. Journal of Biomechanics *45*, 625-633.
37. Ermak, D.L., and Mccammon, J.A. (1978). Brownian Dynamics with Hydrodynamic Interactions. Journal of Chemical Physics *69*, 1352-1360.
38. Eungdamrong, N.J., and Iyengar, R. (2004). Computational approaches for modeling regulatory cellular networks. Trends Cell Biol *14*, 661-669.
39. Faeder, J.R. (2011). Toward a comprehensive language for biological systems. BMC Biol *9*, 68.
40. Fange, D., Berg, O.G., Sjoberg, P., and Elf, J. (2010). Stochastic reaction-diffusion kinetics in the microscopic limit. Proceedings of the National Academy of Sciences of the United States of America *107*, 19820-19825.
41. Fowler, P.W., Hélie, J., Duncan, A., Chavent, M., Koldsø, H., and Sansom, M.S.P. (2016). Membrane stiffness is modified by integral membrane proteins. Soft Matter *12*, 7792-7803.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

42. Fu, Y., Yogurtcu, O.N., Kothari, R., Thorkelsdottir, G., Sodt, A.J., and Johnson, M.E. (2019). An implicit lipid model for efficient reaction-diffusion simulations of protein binding to surfaces of arbitrary topology. J Chem Phys *151*, 124115.
43. Ghaemi, Z., Peterson, J.R., Gruebele, M., and Luthey-Schulten, Z. (2020). An in-silico human cell model reveals the influence of spatial organization on RNA splicing. PLoS Comput Biol *16*, e1007717.
44. Gillespie, D.T. (1976a). General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical-Reactions. Journal of Computational Physics *22*, 403-434.
45. Gillespie, D.T. (1976b). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics *22*, 403-434.
46. Gillespie, D.T. (1977a). Exact stochastic simulation of coupled chemical reactions. J Phys Chem *81*, 2340-2361.
47. Gillespie, D.T. (1977b). Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry *81*, 2340-2361.
48. Greenwald, E.C., Mehta, S., and Zhang, J. (2018). Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks. Chem Rev *118*, 11707-11794.
49. Gregor, T., Wieschaus, E.F., McGregor, A.P., Bialek, W., and Tank, D.W. (2007). Stability and nuclear dynamics of the bicoid morphogen gradient. Cell *130*, 141-152.
50. Grima, R., and Schnell, S. (2008). Modelling reaction kinetics inside cells. Essays Biochem *45*, 41-56.
51. Groen, D., Knap, J., Neumann, P., Suleimenova, D., Veen, L., and Leiter, K. (2019). Mastering the scales: a survey on the benefits of multiscale computing software. Philos Trans A Math Phys Eng Sci *377*, 20180147.
52. Guckenberger, A., and Gekle, S. (2017). Theory and algorithms to compute Helfrich bending forces: a review. Journal of Physics: Condensed Matter *29*, 203001.
53. Gunawardena, J. (2014). Models in biology: 'accurate descriptions of our pathetic thinking'. BMC Biol *12*, 29.
54. Guo, Y., Li, D., Zhang, S., Yang, Y., Liu, J.-J., Wang, X., Liu, C., Milkie, D.E., Moore, R.P., Tulu, U.S.*, et al.* (2018). Visualizing Intracellular Organelle and Cytoskeletal Interactions at Nanoscale Resolution on Millisecond Timescales. Cell *175*, 1430-1442.e1417.
55. Gupta, A., and Mendes, P. (2018). An Overview of Network-Based and -Free Approaches for Stochastic Simulation of Biochemical Systems. Computation (Basel) *6*.
56. Gutenkunst, R.N., Waterfall, J.J., Casey, F.P., Brown, K.S., Myers, C.R., and Sethna, J.P. (2007). Universally Sloppy Parameter Sensitivities in Systems Biology Models. Plos Comput Biol *3*, e189.
57. Halatek, J., and Frey, E. (2012). Highly canalized MinD transfer and MinE sequestration explain the origin of robust MinCDE-protein dynamics. Cell Rep *1*, 741-752.
58. Hallock, M.J., Stone, J.E., Roberts, E., Fry, C., and Luthey-Schulten, Z. (2014). Simulation of reaction diffusion processes over biologically relevant size and time scales using multi-GPU workstations. Parallel Computing *40*, 86-99.
59. Harris, L.A., Hogg, J.S., Tapia, J.J., Sekar, J.A., Gupta, S., Korsunsky, I., Arora, A., Barua, D., Sheehan, R.P., and Faeder, J.R. (2016). BioNetGen 2.2: advances in rule-based modeling. Bioinformatics *32*, 3366-3368.
60. Hasenauer, J., Wolf, V., Kazeroonian, A., and Theis, F.J. (2014). Method of conditional moments (MCM) for the Chemical Master Equation. Journal of Mathematical Biology *69*, 687-735.
61. Health, U.N.I.o. (2018). Enhancing Reproducibility through Rigor and Transparency (grants.nih.gov).
62. Helfrich, W. (1973). Elastic Properties of Lipid Bilayers - Theory and Possible Experiments. Zeitschrift Fur Naturforschung C-a Journal of Biosciences *C 28*, 693-703.
63. Hellander, S., Hellander, A., and Petzold, L. (2012). Reaction-diffusion master equation in the microscopic limit. Physical Review E *85*, 042901.
64. Hellander, S., Hellander, A., and Petzold, L. (2017). Mesoscopic-microscopic spatial stochastic simulation with automatic system partitioning. The Journal of Chemical Physics *147*, 234101.
65. Hodgkin, A.L., and Huxley, A.F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol *117*, 500-544.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

66. Hoffmann, M., Frohner, C., and Noe, F. (2019). ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics. PLoS Comput Biol *15*, e1006830.
67. Howard, M., and Rutenberg, A.D. (2003). Pattern formation inside bacteria: fluctuations due to the low copy number of proteins. Phys Rev Lett *90*, 128102.
68. Isaacson, S.A., and Zhang, Y. (2018). An unstructured mesh convergent reaction-diffusion master equation for reversible reactions. Journal of Computational Physics *374*, 954-983.
69. Janes, K.A., and Lauffenburger, D.A. (2013). Models of signalling networks - what cell biologists can gain from them and give to them. J Cell Sci *126*, 1913-1921.
70. Jilkine, A., and Edelstein-Keshet, L. (2011). A comparison of mathematical models for polarization of single eukaryotic cells in response to guided cues. PLoS Comput Biol *7*, e1001121.
71. Johnson, M.E. (2018). Modeling the Self-Assembly of Protein Complexes through a Rigid-Body Rotational Reaction-Diffusion Algorithm. J Phys Chem B *122*, 11771-11783.
72. Johnson, M.E., and Hummer, G. (2014). Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems. Physical Review X *4*, 031037.
73. Karelina, M., and Kulik, H.J. (2017). Systematic Quantum Mechanical Region Determination in QM/MM Simulation. Journal of Chemical Theory and Computation *13*, 563-576.
74. Karsenti, E., Nédélec, F., and Surrey, T. (2006). Modelling microtubule patterns. Nature Cell Biology *8*, 1204-1211.
75. Keren, K., Pincus, Z., Allen, G.M., Barnhart, E.L., Marriott, G., Mogilner, A., and Theriot, J.A. (2008). Mechanism of shape determination in motile cells. Nature *453*, 475-480.
76. Kerr, R.A., Bartol, T.M., Kaminsky, B., Dittrich, M., Chang, J.C.J., Baden, S.B., Sejnowski, T.J., and Stiles, J.R. (2008). Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces. Siam Journal on Scientific Computing *30*, 3126-3149.
77. Kestler, H.A., Wawra, C., Kracher, B., and Kuhl, M. (2008). Network modeling of signal transduction: establishing the global view. Bioessays *30*, 1110-1125.
78. Kholodenko, B.N., Hoek, J.B., and Westerhoff, H.V. (2000). Why cytoplasmic signalling proteins should be recruited to cell membranes. Trends Cell Biol *10*, 173-178.
79. Kim, J.S., and Yethiraj, A. (2009). Effect of macromolecular crowding on reaction rates: a computational and theoretical study. Biophys J *96*, 1333-1340.
80. Lipkow, K., and Odde, D.J. (2008). Model for Protein Concentration Gradients in the Cytoplasm. Cellular and Molecular Bioengineering *1*, 84-92.
81. Lisman, J.E. (1985). A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. Proc Natl Acad Sci U S A *82*, 3055-3057.
82. Liu, T.L., Upadhyayula, S., Milkie, D.E., Singh, V., Wang, K., Swinburne, I.A., Mosaliganti, K.R., Collins, Z.M., Hiscock, T.W., Shea, J.*, et al.* (2018). Observing the cell in its native state: Imaging subcellular dynamics in multicellular organisms. Science *360*.
83. Loew, L.M., and Schaff, J.C. (2001). The Virtual Cell: a software environment for computational cell biology. Trends Biotechnol *19*, 401-406.
84. Machacek, M., Hodgson, L., Welch, C., Elliott, H., Pertz, O., Nalbant, P., Abell, A., Johnson, G.L., Hahn, K.M., and Danuser, G. (2009). Coordination of Rho GTPase activities during cell protrusion. Nature *461*, 99-103.
85. Majarian, T.D., Murphy, R.F., and Lakdawala, S.S. (2019). Learning the sequence of influenza A genome assembly during viral replication using point process models and fluorescence in situ hybridization. Plos Computational Biology *15*.
86. Matković, K., Gračanin, D., and Hauser, H. (2018). Visual Analytics for Simulation Ensembles (IEEE Press).
87. Mattis, D.C., and Glasser, M.L. (1998). The uses of quantum field theory in diffusion-limited reactions. Rev Mod Phys *70*, 979-1001.
88. Maus, C., Rybacki, S., and Uhrmacher, A.M. (2011). Rule-based multi-level modeling of cell biological systems. BMC Syst Biol *5*, 166.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

89. Michalski, P.J., and Loew, L.M. (2016). SpringSaLaD: A Spatial, Particle-Based Biochemical Simulation Platform with Excluded Volume. Biophys J *110*, 523-529.
90. Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays *35*, 1050-1055.
91. Minton, A.P. (1995). Confinement as a determinant of macromolecular structure and reactivity. II. Effects of weakly attractive interactions between confined macrosolutes and confining structures. Biophys J *68*, 1311-1322.
92. Minton, A.P. (2006). How can biochemical reactions within cells differ from those in test tubes? J Cell Sci *119*, 2863-2869.
93. Moraru, I.I., Schaff, J.C., Slepchenko, B.M., Blinov, M.L., Morgan, F., Lakshminarayana, A., Gao, F., Li, Y., and Loew, L.M. (2008). Virtual Cell modelling and simulation software environment. Iet Systems Biology *2*, 352-362.
94. Morelli, M.J., Wolde, P.R.t., and Allen, R.J. (2009). DNA looping provides stability and robustness to the bacteriophage λ switch. Proceedings of the National Academy of Sciences *106*, 8101-8106.
95. Mullins, R.D., Bieling, P., and Fletcher, D.A. (2018). From solution to surface to filament: actin flux into branched networks. Biophysical Reviews *10*, 1537-1551.
96. Munsky, B., and Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. The Journal of Chemical Physics *124*, 044104.
97. Nedelec, F., and Foethke, D. (2007a). Collective Langevin dynamics of flexible cytoskeletal fibers. New Journal of Physics *9*.
98. Nedelec, F., and Foethke, D. (2007b). Collective Langevin dynamics of flexible cytoskeletal fibers. New Journal of Physics *9*, 427–427.
99. Nickaeen, M., Novak, I.L., Pulford, S., Rumack, A., Brandon, J., Slepchenko, B.M., and Mogilner, A. (2017). A free-boundary model of a motile cell explains turning behavior. PLOS Computational Biology *13*, e1005862.
100. Northrup, S.H., Allison, S.A., and Mccammon, J.A. (1984). Brownian Dynamics Simulation of Diffusion-Influenced Bimolecular Reactions. Journal of Chemical Physics *80*, 1517-1526.
101. Odell, G.M., and Foe, V.E. (2008). An agent-based model contrasts opposite effects of dynamic and stable microtubules on cleavage furrow positioning. Journal of Cell Biology *183*, 471-483.
102. Onsum, M., and Rao, C.V. (2007). A mathematical model for neutrophil gradient sensing and polarization. PLoS Comput Biol *3*, e36.
103. Prinz, F., Schlange, T., and Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? Nature Reviews Drug Discovery *10*, 712-712.
104. Prustel, T., and Meier-Schellersheim, M. (2012). Exact Green's function of the reversible diffusion-influenced reaction for an isolated pair in two dimensions. Journal of Chemical Physics *137*, 054104.
105. Rangamani, P., Levy, M.G., Khan, S., and Oster, G. (2016). Paradoxical signaling regulates structural plasticity in dendritic spines. Proceedings of the National Academy of Sciences *113*, E5298-E5307.
106. Rangamani, P., Mandadap, K.K., and Oster, G. (2014). Protein-induced membrane curvature alters local membrane tension. Biophys J *107*, 751-762.
107. Raskin, D.M., and de Boer, P.A.J. (1999). Rapid pole-to-pole oscillation of a protein required for directing division to the middle of Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *96*, 4971-4976.
108. reproducciblebiomodels.org (2020). Center for Reproducible Biomedical Modeling.
109. Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics *16*, 276-277.
110. Richmond, B.G., Wright, B.W., Grosse, L., Dechow, P.C., Ross, C.F., Spencer, M.A., and Strait, D.S. (2005). Finite element analysis in functional morphology. Anatomical Record Part a-Discoveries in Molecular Cellular and Evolutionary Biology *283a*, 259-274.
111. Rubinstein, B., Larripa, K., Sommi, P., and Mogilner, A. (2009). The elasticity of motor–microtubule bundles and shape of the mitotic spindle. Physical Biology *6*, 016005.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

112. Sanghvi, J.C., Regot, S., Carrasco, S., Karr, J.R., Gutschow, M.V., Bolival, B., Jr., and Covert, M.W. (2013). Accelerated discovery via a whole-cell model. Nat Methods *10*, 1192-1195.
113. SBML.org (2020). The Systems Biology Markup Language (SBML.org).
114. Schaff, J.C., Slepchenko, B.M., and Loew, L.M. (2000). Physiological modeling with virtual cell framework. Numerical Computer Methods, Part C *321*, 1-23.
115. Schaff, J.C., Vasilescu, D., Moraru, II, Loew, L.M., and Blinov, M.L. (2016). Rule-based modeling with Virtual Cell. Bioinformatics *32*, 2880-2882.
116. Schillings, C., Sunnåker, M., Stelling, J., and Schwab, C. (2015). Efficient Characterization of Parametric Uncertainty of Complex (Bio)chemical Networks. Plos Comput Biol *11*, e1004457.
117. Schnoerr, D., Sanguinetti, G., and Grima, R. (2017). Approximation and inference methods for stochastic biochemical kinetics-a tutorial review. Journal of Physics a-Mathematical and Theoretical *50*, 1-60.
118. Schoneberg, J., and Noe, F. (2013). ReaDDy - A Software for Particle-Based Reaction-Diffusion Dynamics in Crowded Cellular Environments. Plos One *8*.
119. Schoneberg, J., Ullrich, A., and Noe, F. (2014). Simulation tools for particle-based reaction-diffusion dynamics in continuous space. Bmc Biophysics *7*.
120. Schreiber, G., Haran, G., and Zhou, H.X. (2009). Fundamental Aspects of Protein-Protein Association Kinetics. Chemical Reviews *109*, 839-860.
121. Sekar, J.A.P., Tapia, J.J., and Faeder, J.R. (2017). Automated visualization of rule-based models. PLoS Comput Biol *13*, e1005857.
122. Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y.*, et al.* (2010). Atomic-Level Characterization of the Structural Dynamics of Proteins. Science *330*, 341-346.
123. Shlemov, A., Golyandina, N., Holloway, D., and Spirov, A. (2015). Shaped Singular Spectrum Analysis for Quantifying Gene Expression, with Application to the Early Drosophila Embryo. In BioMed Research International.
124. Smith, A.M., Xu, W., Sun, Y., Faeder, J.R., and Marai, G.E. (2012). RuleBender: integrated modeling, simulation and visualization for rule-based intracellular biochemistry. BMC Bioinformatics *13 Suppl 8*, S3.
125. Smith, S., and Grima, R. (2016). Breakdown of the reaction-diffusion master equation with nonelementary rates. Phys Rev E *93*, 052135.
126. Snow, C.D., Nguyen, H., Pande, V.S., and Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. Nature *420*, 102-106.
127. Sokolowski, T.R., Paijmans, J., Bossen, L., Miedema, T., Wehrens, M., Becker, N.B., Kaizu, K., Takahashi, K., Dogterom, M., and Ten Wolde, P.R. (2019). eGFRD in all dimensions. J Chem Phys *150*, 054108.
128. Spagnolo, J.F., Rossignol, E., Bullitt, E., and Kirkegaard, K. (2010). Enzymatic and nonenzymatic functions of viral RNA-dependent RNA polymerases within oligomeric arrays. RNA *16*, 382-393.
129. Szabo, A., Schulten, K., and Schulten, Z. (1980). 1st Passage Time Approach to Diffusion Controlled Reactions. Journal of Chemical Physics *72*, 4350-4357.
130. Takahashi, K., Arjunan, S.N., and Tomita, M. (2005). Space in systems biology of signaling pathways--towards intracellular molecular crowding in silico. FEBS Lett *579*, 1783-1788.
131. Takahashi, K., Tanase-Nicola, S., and ten Wolde, P.R. (2010). Spatio-temporal correlations can drastically change the response of a MAPK pathway. Proceedings of the National Academy of Sciences of the United States of America *107*, 2473-2478.
132. Tanaka, S., Sichau, D., and Iber, D. (2015). LBIBCell: a cell-based simulation environment for morphogenetic problems. Bioinformatics *31*, 2340-2347.
133. Tauber, U.C., Howard, M., and Vollmayr-Lee, B.P. (2005). Applications of field-theoretic renormalization group methods to reaction-diffusion problems. J Phys A-Math Gen *38*, R79-R131.
134. Tay, S., Hughey, J.J., Lee, T.K., Lipniacki, T., Quake, S.R., and Covert, M.W. (2010). Single-cell NF-κB dynamics reveal digital activation and analogue information processing. Nature *466*, 267-271.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.02.185595; this version posted July 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

135. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols *7*, 562-578.
136. Turing, A.M. (1952). The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences *237*, 37-72.
137. Tyson, J.J., Chen, K.C., and Novak, B. (2003). Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. Curr Opin Cell Biol *15*, 221-231.
138. Van Kampen, N.G. (2007). Stochastic Processes in Physics and Chemistry, Third edn (Amsterdam: Elsevier).
139. van Zon, J.S., and ten Wolde, P.R. (2005). Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. Physical Review Letters *94*.
140. Varga, M., Fu, Y., Loggia, S., Yogurtcu, O.N., and Johnson, M.E. (2020). NERDSS: a nonequilibrium simulator for multibody self-assembly at the cellular scale. Biophysical Journal *118*, P3026-3040.
141. Vilar, J.M., Kueh, H.Y., Barkai, N., and Leibler, S. (2002). Mechanisms of noise-resistance in genetic oscillators. Proc Natl Acad Sci U S A *99*, 5988-5992.
142. von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M. (2000). The segment polarity network is a robust developmental module. Nature *406*, 188-192.
143. von Smoluchowski, M. (1917). Attempt to derive a mathematical theory of coagulation kinetics in colloidal solutions. Z Phys Chem *92*, 129.
144. Wedlich-Soldner, R., Altschuler, S., Wu, L., and Li, R. (2003). Spontaneous cell polarization through actomyosin-based delivery of the Cdc42 GTPase. Science *299*, 1231-1235.
145. Wolf, V., Goel, R., Mateescu, M., and Henzinger, T.A. (2010). Solving the chemical master equation using sliding windows. BMC Systems Biology *4*, 42.
146. Wu, Y., Vendome, J., Shapiro, L., Ben-Shaul, A., and Honig, B. (2011). Transforming binding affinities from three dimensions to two with application to cadherin clustering. Nature *475*, 510-513.
147. Wu, Z., Su, M., Tong, C., Wu, M., and Liu, J. (2018). Membrane shape-mediated wave propagation of cortical protein dynamics. Nature Communications *9*, 136.
148. Yasunaga, A., Murad, Y., and Li, I.T.S. (2019). Quantifying molecular tension—classifications, interpretations and limitations of force sensors. Physical Biology *17*, 011001.
149. Yogurtcu, O.N., and Johnson, M.E. (2015). Theory of bi-molecular association dynamics in 2D for accurate model and experimental parameterization of binding rates. J Chem Phys *143*, 084117.
150. Yogurtcu, O.N., and Johnson, M.E. (2018). Cytoplasmic proteins can exploit membrane localization to trigger functional assembly. PLoS Comp Biol *14*, e1006031.
151. Zhang, F., Angermann, B.R., and Meier-Schellersheim, M. (2013). The Simmune Modeler visual interface for creating signaling networks based on bi-molecular interactions. Bioinformatics *29*, 1229-1230.
152. Zhou, H.X., Rivas, G., and Minton, A.P. (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. Annu Rev Biophys *37*, 375-397.
153. Zimmerman, S.B., and Minton, A.P. (1993). Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct *22*, 27-65.
154. Zuckerman, D.M. (2011). Equilibrium Sampling in Biomolecular Simulations. Annual Review of Biophysics *40*, 41-62.