Lost in presumption: stochastic reactions in spatial models

Anel Mahmutovic, David Fange, Otto G Berg & Johan Elf

Physical modeling is increasingly important for generating insights into intracellular processes. We describe situations in which combined spatial and stochastic aspects of chemical reactions are needed to capture the relevant dynamics of biochemical systems.

Consider a situation in which intracellular spatial oscillations are observed for a protein of interest. Some of the protein's biochemical interactions and their parameters have been characterized, so it is straightforward to draw a cartoon showing how the components may interact. However, on the basis of the cartoon alone, it is very hard to discern whether the components and their interactions are sufficient and necessary to explain the observed oscillation phenomenon. The reason is that the intracellular reactions are nonlinear, noisy, out of equilibrium and dependent on local concentrations, which can make the outcome of intracellular interactions highly nonintuitive. One way to resolve this problem is to make a quantitative description—that is, a mathematical model—of the cartoon. A well-formulated model makes it possible to identify whether the reactions in the cartoon are sufficient to give rise to the observed phenotype and, if so, under which conditions. Building such a mathematical model is not straightforward, however, especially because it is rarely obvious a priori which level of physical detail is required to capture the biochemical behavior of the cellular process.

At a high level of detail, a framework based on molecular dynamics keeps track of the positions of atoms in molecules and the forces between them. At a lower level of detail, particle-based frameworks keep track of the positions of individual molecules and when they react. Both of these approaches in many cases contain more detail than is required to understand a particular biological process; even more coarse-grained frameworks are therefore commonly used.

Here we describe commonly used physical frameworks for modeling biochemical systems at different levels of detail (**Fig. 1**). In particular, we focus on the specific situations requiring frameworks that consider both the positions of individual molecules and the stochastic nature of the chemical reactions between them.

Well-stirred deterministic models

If the molecules of interest diffuse quickly and, on average, have time to move through the cell between reactions, it is safe to assume that it is only the total number of molecules of each species, and not their subcellular location, that determines the overall reaction rates. If, in addition, there are enough molecules of all species involved such that stochastic copy-number fluctuations play a minor role, it is possible to model how the average concentrations change in time without considering molecular copy-number distributions. The dynamics of the average concentrations are expressed in ordinary differential equations (ODEs), often referred to as reaction rate equations. This level of modeling is a natural reference for more detailed models also when the conditions for its validity are not strictly met.

Low-copy-number fluctuation models

When the number of molecules of some species is low or the relaxation to steady state is slow[1], the stochastic nature of chemical reactions may influence the dynamics of the chemical reaction system. Consequently, the average number of molecules can no longer

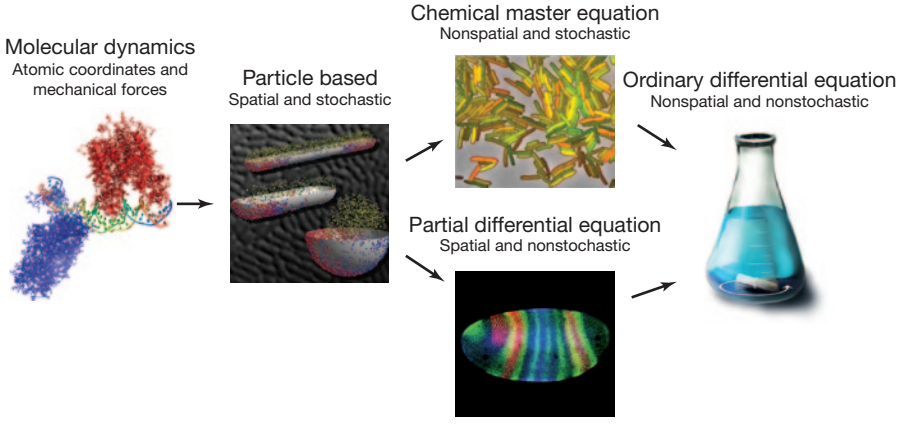

Figure 1 | Different levels of quantitative modeling frameworks for intracellular chemistry. Microscopy images and graphics courtesy of M. Elowitz, S. Paddock, S.B. Carroll and I. Barkefors.

Anel Mahmutovic, David Fange, Otto G. Berg and Johan Elf are in the Department of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Sweden.
e-mail: johan.elf@icm.uu.se

be described by deterministic reaction rate equations. This is because the average rate of change in species concentration is typically not equal to the rate of change evaluated for the average concentration: that is, for concentration x and reaction rate $f(x)$, $\langle f(x)\rangle \neq f(\langle x\rangle)$. The modeling framework that accounts for the stochastic aspects of chemical reactions, albeit without spatial consideration, is the chemical master equation (CME)[2]. In contrast to the deterministic rate equations used in ODE models, the CME describes the probability distribution of molecule copy numbers and how this distribution changes over time. The master equation can, for example, be used to model stochastic gene expression in a single cell. In this case, a few genes per cell make mRNA molecules that can be translated many times and thus give rise to stochastic bursts in the number of proteins[3]. The CME describes how parameters such as transcription rates, mRNA lifetimes and translation rates contribute to the observed cell-to-cell variability in protein expression[4].

Deterministic spatial models

When the cell is large in relation to the average distance that reactants diffuse in between reactions, the reaction rates will depend on local concentrations. If, at the same time, each reactant has a large number of reaction partners within its diffusion range, local concentration averages describe the state of the system accurately, and it is not necessary to keep track of individual molecules. A model that captures this scenario must describe how concentration averages vary in different parts of a cell or organism and how these local concentration averages change over time. These quantitative models are formulated in coupled partial differential equations (PDEs)[5], which have a rich history in describing biological pattern formation[6].

Spatial stochastic models

Although the total number of reactant molecules in the cell may be sufficiently high to motivate an ODE model if the molecules were well-stirred, the number of reaction partners within diffusion range of a reactant can be low and subject to stochastic fluctuations. In such a case, particle-based modeling frameworks that consider both the spatial and stochastic aspects of chemical reactions may be important. There are no general rules for when this detail of modeling is needed to describe the system at a desired level of accuracy. We will instead exemplify, with three different situations, where a particle-based framework gives radically different results compared to that of a nonspatial stochastic treatment (that is, using CME) or a spatial nonstochastic treatment (using PDE).

Implementing spatial stochastic models can be finicky; there are several useful free software tools available for this purpose, such as GFRD, MesoRD and Smoldyn[7–9].

Example 1: nonlinear responses to uneven concentrations. Consider a substance S synthesized by sparsely distributed enzyme molecules, X, that are diffusing slowly (**Fig. 2a**). S is consumed by enzyme molecules E via a conventional Michaelis-Menten reaction scheme. The S molecules do not have time to mix in the entire cell before consumption; this leads to concentration peaks around each synthesis enzyme, which randomly changes location because of diffusion. As it is possible to locally saturate the consumption enzymes, E, the overall rate of S degradation is not as fast as it would be if S were evenly spread in the system. Particle-based models capture this behavior and show a higher overall concentration of S than the nonspatial stochastic description (CME) and the spatial,

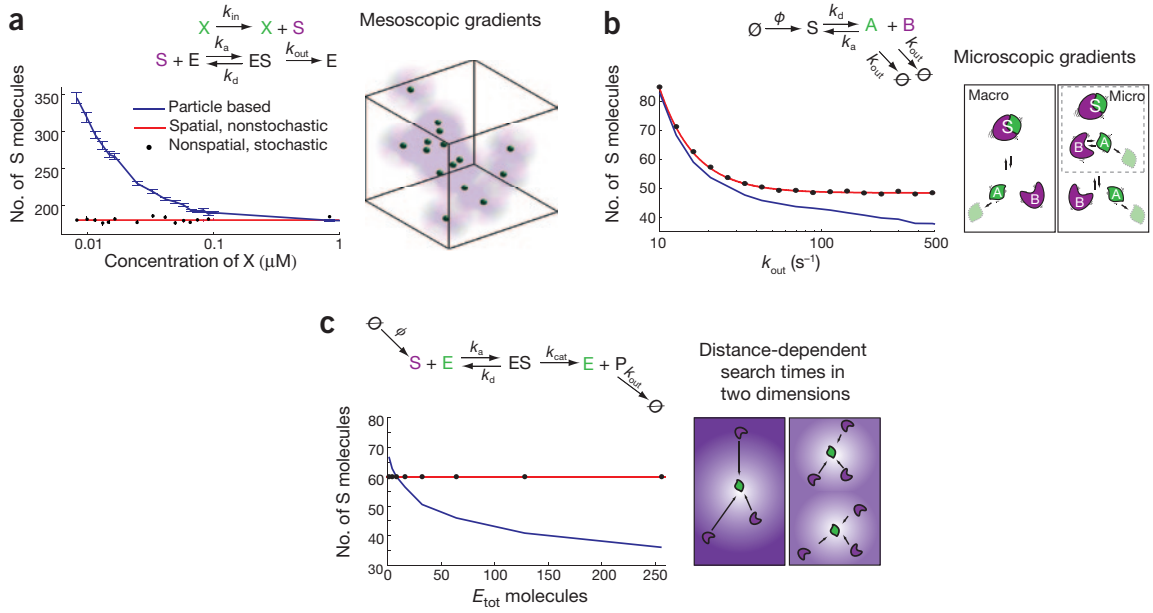

Figure 2 | Examples of combined spatial and stochastic effects in three simple systems. (**a**) Nonlinear responses to uneven concentrations. Right, illustration of locally produced and slowly diffusing molecules (S) that locally saturate the degradation enzymes (E). Left, reaction scheme modeled with the indicated frameworks. In the graph, the total S synthesis rate is held constant for varying synthesizing enzyme (X) concentration by a changing substrate synthesis rate per enzyme (k_{in}). k_a, association; k_d, dissociation; and k_{out}, degradation rate constants. (**b**) Interference at molecular length scales. Right, illustration of the difference in available free A and B molecules when binding and release are treated macroscopically versus microscopically. Left, reaction scheme modeled with the indicated frameworks. ϕ, influx rate. (**c**) Long-range correlations on membranes. Right, illustration of the decrease in search time when the enzyme concentration doubles. Left, reaction scheme modeled with the indicated frameworks. The influx rate increases in proportion to the total number of degradation enzymes, E_{tot}. See the **Supplementary Note** for parameters, simulation details and analytical results.

nonstochastic description (PDE). This situation is exemplified *in vivo* by local mRNA synthesis near a relatively stationary gene in a prokaryotic cell. Local peaks in mRNA concentration near the transcription sites have recently been observed in *Escherichia coli*[10], suggesting that the observed mRNAs do not diffuse through the cell. These mRNA molecules may saturate the local ribosomes such that the overall rate of translation is slower than it would be if the same mRNA molecules were spread evenly over the cell.

The situation we describe here (sparsely distributed synthesis molecules) is but one example of how stochastic variations in the spatial distribution of a species S can arise. What is more important is how the uneven distribution of S molecules influences the average rate of their downstream reactions. If the S molecules contribute linearly to the rate of downstream reactions, the uneven distribution will not affect the spatially averaged rate. However, if the S molecules contribute nonlinearly to the rate, as for saturable enzyme reactions, the spatially averaged rate may depend on their distribution. In the extreme case of reaction schemes with sharp activation thresholds, such as bistable or excitable systems, spatially heterogeneous fluctuations may even drive the system into alternative steady states because the fluctuations can make the system pass the threshold in some part of the cell[11,12].

Example 2: interference at molecular length scales. An intriguing situation arises when reaction rates are close to diffusion limited such that newly dissociated reactants have a high probability of reassociating before diffusing apart. This implies that dissociation processes include multiple rebinding events before the molecules actually lose spatial correlation[13]. In a CME or PDE treatment, these multiple rounds of reassociations are included in the dissociation rate constant, which therefore includes both the actual dissociation event and the diffusive transport to a position where the molecules are uncorrelated. This approximation fails when one of the newly dissociated molecules can enter or influence other processes before the molecules have lost spatial correlation.

Consider S molecules introduced at random places in the cell (**Fig. 2b**), after which S can dissociate into A and B. Subsequently the A and B molecules can either reform S or enter into another process, which is here exemplified by a degradation process. When degradation is rapid, A and B will degrade before they have had time to reach uncorrelated positions after dissociation. Here a particle-based method, which correctly captures that A and B molecule are available for degradation before they have reached uncorrelated positions, displays a lower steady-state concentration than do the CME and PDE, which assume that the A and B molecules are bound to each other until they have moved far apart.

This situation would be uncommon for chemical reactions in which the reactants are freely diffusing in three dimensions. Here reactants lose spatial correlation at a length of only a few molecular radii, and few competing reactions can occur before the molecules reach this distance. However, in the case where the same enzyme can modify a macromolecule at several positions before diffusing away[14], the competing reactant is the same molecule that was involved in the previous step. Similar situations will arise in scaffolds for signaling pathways or in receptor clusters[15].

Example 3: long-range correlations on membranes. Reactions between molecules on a membrane—that is, in two dimensions—are radically different from those in three dimensions. In two dimensions, the absolute distances between molecules will influence their reaction rates at all length scales. For instance: in three dimensions, a molecule in a large reaction volume has the same probability of first colliding with a molecule that is 10 molecule radii away as with one that is 100 radii away. On a two-dimensional (2D) membrane, the molecule is three times more likely to first collide with the molecule that is ten radii away. As a consequence, the reaction-rate constants in a 2D system can be strictly defined only microscopically, with the distance dependence explicitly accounted for. Modeling frameworks based on macroscopic rates assume invariance to absolute distances and are thus not well-defined in two dimensions. Put differently, in two dimensions, the macroscopic reaction rate 'constants' are dependent on time and/or concentration, which may result in inconsistent results unless concentration variations are kept small.

To illustrate these points, consider a system in which all species are diffusing on a flat 2D surface (**Fig. 2c**). Enzyme molecules can convert substrate molecules into product via a Michaelis-Menten–type reaction scheme. New substrate molecules are introduced at random positions on the surface, and the product is degraded such that the substrate flow is kept out of equilibrium. We show the effect of absolute distances in 2D systems by gradually increasing the substrate influx while proportionally increasing the number of enzyme molecules. In a corresponding 3D system, the same fractional increase, α, of both substrate influx, αk_{in}, and enzyme concentration, α[E], would give an unchanged product flux per enzyme $k_{in}/[E] = k_{cat}/(1 + K_M/[S])$, implying that the substrate concentration, [S], remains constant. However, in this 2D system, the substrate molecules are more likely to react with the closest enzyme, which means that an enzyme density increase leads to shorter substrate search times and, therefore, a higher enzyme turnover. Thus, in the 2D system using a particle-based framework, substrate concentration decreases with increasing enzyme density. See the **Supplementary Note** for the analytical treatment and solution of this system. The situation in one dimension, such as for proteins sliding on DNA or filaments, is even more constrained because molecules cannot bypass each other[16].

Why make it so complicated?

Is it necessary to make complicated models that include both noise and space when there are usually insufficient *in vivo* data to fit the parameters even in an ordinary reaction-rate ODE model? In some cases, it is obvious that spatial stochastic models are needed: for example, in trying to understand why a protein binding pattern nucleates at random positions in space[11] or how an embryo minimizes noise in a segmentation gradient[17]. In other cases, particle-based models are desired because a less detailed modeling framework would in fact require a more complex model to explain the data. For instance, in our third example above, cooperative substrate binding would be needed to account for the decrease in substrate concentration unless the peculiarities of 2D kinetics are accounted for. A final reason for using a detailed framework is that it makes it possible to test whether less detailed models are still physically valid when geometrical constraints and physical limitation on

microscopic reaction-diffusion couplings are considered.

In summary, in specific situations, the properties of intracellular processes will depend on the spatial as well as stochastic aspects of physical chemistry. This can be attributed to mesoscopic gradients arising from spatial variations in molecular copy numbers and slow mixing (example 1) or to microscopic gradients arising when diffusion-controlled reactions are perturbed from steady state (example 2). Such situations are natural parts of intracellular chemistry, and recognizing the consequences on the operation of a biological system may be as important as identifying all the molecular components. At the same time, combined spatial and stochastic considerations should obviously be introduced only when it makes it easier to understand the biological phenomena of interest.

Note: Supplementary information is available at http://www.nature.com/doifinder/10.1038/nmeth.2253.

ACKNOWLEDGMENTS

This work was supported by the Swedish Research Council (VR), Knut and Alice Wallenberg Foundation (KAW) and European Research Council (ERC). We are grateful for comments on the text from I. Barkefors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Elf, J., Paulsson, J., Berg, O.G. & Ehrenberg, M. *Biophys. J.* **84**, 154–170 (2003).
2. Van Kampen, N.G. *Stochastic Processes in Physics and Chemistry* 3rd edn. doi:10.1016/j.bbr.2011.03.031 (North Holland, Amsterdam, 2007).
3. Paulsson, J. *Nature* **427**, 415–418 (2004).
4. Berg, O.G. *J. Theor. Biol.* **71**, 587–603 (1978).
5. Angermann, B.R. *et al. Nat. Methods* **9**, 283–289 (2012).
6. Turing, A.M. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **237**, 37–72 (1952).
7. van Zon, J.S. & ten Wolde, P.R. *J. Chem. Phys.* **123**, 234910 (2005).
8. Fange, D., Mahmutovic, A. & Elf, J. *Bioinformatics* published online, doi:10.1093/bioinformatics/bts584 (7 October 2012).
9. Andrews, S.S., Addy, N.J., Brent, R. & Arkin, A.P. *PLoS Comput. Biol.* **6**, e1000705 (2010).
10. Montero Llopis, P. *et al. Nature* **466**, 77–81 (2010).
11. Fange, D. & Elf, J. *PLoS Comput. Biol.* **2**, e80 (2006).
12. Elf, J. & Ehrenberg, M. *Syst. Biol. (Stevenage)* **1**, 230–236 (2004).
13. Berg, O.G. *Chem. Phys.* **31**, 47–57 (1978).
14. Takahashi, K., Tanase-Nicola, S. & ten Wolde, P.R. *Proc. Natl. Acad. Sci. USA* **107**, 2473–2478 (2010).
15. Oh, D. *et al. Proc. Natl. Acad. Sci. USA* **109**, 14024–14029 (2012).
16. Li, G.-W., Berg, O.G. & Elf, J. *Nat. Phys.* **5**, 294–297 (2009).
17. Howard, M. & ten Wolde, P.R. *Phys. Rev. Lett.* **95**, 208103 (2005).