THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Spatial Stochastic Intracellular Kinetics

Citation for published version:
Smith, S & Grima, R 2018, 'Spatial Stochastic Intracellular Kinetics: A Review of Modelling Approaches',
Bulletin of Mathematical Biology. https://doi.org/10.1007/s11538-018-0443-1

Digital Object Identifier (DOI):
10.1007/s11538-018-0443-1

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Bulletin of Mathematical Biology

OPEN ACCESS

Society for
Mathematical
Biology

CrossMark

Spatial Stochastic Intracellular Kinetics: A Review of Modelling Approaches

Stephen Smith[1] · **Ramon Grima**[1]

Abstract Models of chemical kinetics that incorporate both stochasticity and diffusion are an increasingly common tool for studying biology. The variety of competing models is vast, but two stand out by virtue of their popularity: the reaction–diffusion master equation and Brownian dynamics. In this review, we critically address a number of open questions surrounding these models: How can they be justified physically? How do they relate to each other? How do they fit into the wider landscape of chemical models, ranging from the rate equations to molecular dynamics? This review assumes no prior knowledge of modelling chemical kinetics and should be accessible to a wide range of readers.

1 Introduction

Chemical reactions are the building blocks of biology. Substrates bind to enzymes, messenger RNA binds to ribosomes, proteins bind to DNA, and amino acids bind to each other—the cumulative effect is ultimately life as we know it. It is perhaps unintuitive to think in these terms, but any observable behaviour of a living organism (e.g. a human reading this sentence) can be understood as a series of interactions between molecules. The question of how microscopic chemistry gives rise to macroscopic biology is one of the greatest open scientific problems. Modelling is key to answering this question: our understanding of the underlying chemistry is good, as

✉ Ramon Grima
Ramon.Grima@ed.ac.uk

1 School of Biological Sciences, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JR, Scotland, UK

🙂 Springer

is our knowledge of the observable biology; what is missing is a model linking the distinct scales together.

A system can be modelled in a great variety of ways, but not all models are equally useful. It is theoretically possible to model the biochemistry in an entire cell using Schrödinger's equation, but it would not be appropriate. The challenge is to select a model incorporating the salient details of a system while leaving out the extraneous ones.

For instance, consider a simple molecule like carbon dioxide, CO_2. At the coarsest level, we could model the *concentration* of CO_2, which refers to the total number of carbon dioxide molecules divided by the volume of whatever container the molecules are in (such as a cell or a test tube). The numerical value which we assign to the concentration will depend on the concentrations of other chemical species in the container and the rates of any reactions which involve these species. For instance, there might be some molecules of carbon monoxide, CO, and some molecules of oxygen, O, and there might be a reaction of the form:

$$CO + O \xrightarrow{k} CO_2, \tag{1}$$

where k denotes the rate at which CO and O are converted into CO_2. The basic model which describes chemical kinetics at this level of complexity is the rate equations (REs) (Chen et al. 2010).

Alternatively, at a considerably more complex level we could model each atom of each molecule individually, by considering the forces exerted by each atom in the volume on every other atom. The three atoms which make up any single CO_2 molecule are bonded together and so will exert very strong forces on each other (these forces are traditionally modelled as springs), but they will also experience weaker forces from the rest of the atoms (e.g. electrostatic forces). This is a more complex model because if our container contains, say, 1000 C atoms and 2000 O atoms we will need to keep track of 9000 distinct quantities (the locations of each atom in 3 spatial dimensions) rather than 3 in the REs (the concentrations of each species). The basic model which describes chemical kinetics at this level of complexity is molecular dynamics (MD) (Frenkel 2001).

There are a huge variety of models spanning the range of complexity from the REs to MD. We can imagine these models as points on a "complexity scale", ranging from the coarsest (the REs) to the extremely complex (the Schrödinger equation), as shown in Fig. 1. On the left, we find the REs. As we move up the complexity scale, we gain microscopic detail at the cost of more difficult mathematics or longer computation times, ultimately resulting in MD. For example, one of the main assumptions behind the REs is that the diffusion coefficients of all molecules are infinitely fast—relaxing this assumption pushes us up the complexity scale to a model known as the reaction–diffusion equations (RDEs) (Murray 2001), which is always going to be more accurate than the REs, but correspondingly will always incur a greater computational cost.

The two models which we focus on in this review are at an intermediate level of complexity: the reaction–diffusion master equation (RDME) (Gillespie et al. 2014) and Brownian dynamics (BD) (Lipkov et al. 2011). As we will go on to show, these models are very different, though they both model the same kinds of processes at roughly
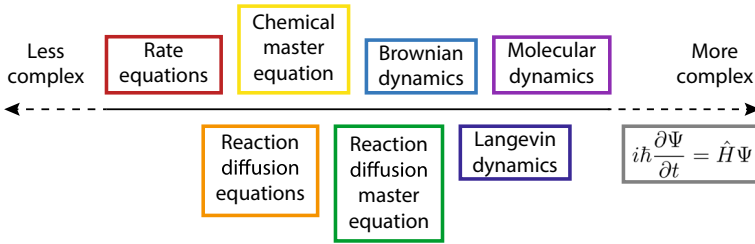
Fig. 1 The scale of model complexity, ranging from coarse to extremely complex. Between the extremes are the models of interest in chemical kinetics, ranging from the rate equations (red) to molecular dynamics (violet). Reviews of the models *not* studied in depth in this article can be found in Refs. Murray (2001), Gillespie et al. (2013) and Durrant (2011)

the same level of detail. Unlike the REs, the RDME and BD both acknowledge that molecules are discrete entities, but they do not go down to atomic-level resolution like MD. The RDME and BD do not explicitly model water molecules (unlike MD), but they do model the *effect* of water molecules (diffusion) which is beyond the scope of the REs.

The reason for focussing on the RDME and BD is that they are likely to be the appropriate level of modelling detail for understanding the link between microscopic chemistry and macroscopic biology (Klann and Koeppl 2012). They are the simplest models which explicitly model individual reactions and the diffusion of individual molecules, and so are not too distant from chemistry as we understand it intuitively. Yet they are not too computationally intensive: both models have already been used to study systems on the scale of an entire cell (Fange and Elf 2006; Sturrock et al. 2013; Lipkow et al. 2005; Andrews et al. 2010), and it is surely only a matter of time until multicellular organisms are within their scope.

It is worth mentioning at this point that this review will not go into detail about the computational methods used to implement these models in practice (the so-called simulation algorithms). A relatively recent and comprehensive review of this issue can be found in Ref. Erban et al. (2007). Instead, our concern will be with the validity of the RDME and BD, how they are related, when it is appropriate to use them, and how they fit into the modelling complexity scale. Atypically, the subject of our studies will be the models themselves rather than the systems to which they could be applied.

Our review is split into three sections. In Sect. 2 we discuss the mathematical origins of the RDME, followed by a detailed analysis of how it relates to other models on the complexity scale. In Sect. 3 we discuss the physical origins of BD, followed by a rigorous discussion of its position in the complexity scale. In Sect. 4 we briefly discuss the extent to which BD and the RDME agree with each other. We conclude with a discussion in Sect. 5.

2 The Reaction–Diffusion Master Equation

Consider the following chemical reaction system:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C. \tag{2}$$

This denotes that a molecule of type A can bind to a molecule of type B to create a molecule of type C and that this process occurs with rate k_1. Furthermore, a molecule of type C can spontaneously unbind into one molecule of type A and one molecule of type B, a process which occurs with rate k_2. As mentioned in Introduction, the simplest method of modelling systems like (2) is to describe the concentrations (number of molecules per unit volume) of A, B and C as differentiable functions of time, denoted $[A]$, $[B]$ and $[C]$, respectively. These functions are implicitly defined as the solutions to a set of ordinary differential equations, the REs:

$$
\begin{aligned}
\frac{d[A]}{dt} &= -k_1[A][B] + k_2[C], \\
\frac{d[B]}{dt} &= -k_1[A][B] + k_2[C], \\
\frac{d[C]}{dt} &= k_1[A][B] - k_2[C].
\end{aligned}
\tag{3}
$$

The equation for $\frac{d[A]}{dt}$, for example, states that the rate of change of $[A]$ is equal to minus the rate at which A molecules are removed from the system (the rate of the binding reaction, k_1, multiplied by the product of concentrations of the reactants, $[A][B]$) plus the rate at which A molecules are added to the system (the rate of the unbinding reaction, k_2, multiplied by the concentration of the reactant, $[C]$). The functional form of these rates follows from the principle of mass action kinetics (Erdi and Janos 1989), which states that the rate of a reaction is proportional to the product of concentrations of the reactants.

Equation (3), coupled with initial concentrations of each species, provides all the information required to ascertain the concentrations of A, B and C at any time in the future, either by using a numerical ODE solver (Butcher 2016), or by analytically solving the equations.

The REs are simple to derive for any system, simple to understand, and easy to solve with a computer, and therefore they remain by far the most common model of chemical kinetics in use today, in fields ranging from physical chemistry (Rickard 1997) to cell biology (Schnell and Mendoza 1997).

However, the RE model relies heavily on the assumption that concentrations are differentiable functions of time, which is clearly untrue since numbers of molecules must be integer-valued, and therefore discrete. This assumption becomes particularly egregious when the concentrations are small, so that a system may contain only a few tens or hundreds of molecules of each species. In addition to this, it must be noted that chemical kinetics is inherently probabilistic, for a variety of reasons. For one, chemical reactions are quantum mechanical events (Atkins et al. 2018), though the details of this are beyond the scope of this article. Another reason is that bimolecular reactions occur only when the two reactants diffuse close enough together to react, and diffusion (the cumulative effect of huge numbers of collisions with water molecules) is such a complex process that it is typically modelled as random (Gillespie and Seitaridou 2012).

Thinking along the lines of discreteness and randomness leads us to consider not concentrations $[A]$, $[B]$ and $[C]$, but rather the joint probability mass function

$P(n_A, n_B, n_C; t)$, the probability that the system contains exactly n_A molecules of A, n_B molecules of B and n_C molecules of C at time t. Though this quantity may seem hopelessly complicated, it turns out that we can say quite a lot about it.

Consider a very short time period Δt, so short that at most one reaction can happen in it, then consider what we can say about $P(n_A, n_B, n_C; t + \Delta t)$ in terms of $P(n_A, n_B, n_C; t)$. We can end up with exactly n_A, n_B, n_C molecules at time $t + \Delta t$ in three different ways: (I) if there were $n_A + 1, n_B + 1, n_C - 1$ molecules at time t and a binding reaction happened in the interval $[t, t + \Delta t)$; (II) if there were $n_A - 1, n_B - 1, n_C + 1$ molecules at time t and an unbinding reaction happened in $[t, t + \Delta t)$; and (III) if there were n_A, n_B, n_C molecules at time t and no reactions happened in $[t, t + \Delta t)$. So we can write:

$$P(n_A, n_B, n_C; t + \Delta t) = P(\mathrm{I}) + P(\mathrm{II}) + P(\mathrm{III}), \tag{4}$$

where $P(i)$ represents the probability that scenario (i) happened. By the definition of conditional probability (Loeve 1977), we can write the following for $P(\mathrm{I})$:

$$P(I) = P(n_A + 1, n_B + 1,$$
$$n_C - 1 \text{ molecules at time } t \text{ and a binding reaction happened}) \tag{5}$$
$$= P(n_A + 1, n_B + 1, n_C - 1; t)$$
$$\times P(\text{a binding reaction happened in time } \Delta t \mid n_A + 1,$$
$$n_B + 1, n_C - 1 \text{ molecules at time } t), \tag{6}$$

where the conditional symbol | means "given that".

At this point in the analysis we need to make an assumption, namely that the waiting times between chemical reactions are exponentially distributed. For unimolecular reactions (i.e. unbinding), Fermi's golden rule implies that waiting times are very close to exponential (Dirac 1927; Fermi 1950). For bimolecular reactions (i.e. binding) the reality is more complicated: reacting molecules must diffuse close together, then collide with a sufficiently high energy, at the correct orientation (Atkins et al. 2018). Although the underlying processes involved here are ultimately deterministic (at least at the level we go to in this article), they are so complex that they appear to be random, and the overall waiting time will appear to follow some probability distribution. The exponential distribution is chosen because of the assumption of *memorylessness*, meaning that the future of the system depends only on the current state (i.e. molecule numbers) and not on the states which preceded it. Exponential waiting times are a direct consequence of this assumption. Though it is a straightforward assumption, it is not strictly true, for instance, imagine that a system is currently in the state $(n_A = 10, \ n_B = 10, \ n_C = 10)$. According to the memorylessness assumption, the probability that a binding reaction happens next is independent of how the system got into its current state, whether from state $(n_A = 11, \ n_B = 11, \ n_C = 9)$ and a binding reaction, or from state $(n_A = 9, \ n_B = 9, \ n_C = 11)$ and an unbinding reaction. Yet if an unbinding reaction happened very recently, then we know that the products of that reaction will be close together, and so will be significantly more likely to bind than a typical pair of reactants. In other words, binding reactions are more likely in

the immediate aftermath of an unbinding, violating the memorylessness assumption. Choosing the exponential distribution essentially amounts to ignoring this rather subtle effect, but in its favour the exponential distribution has very useful mathematical properties.

One of the nice mathematical properties of exponentially distributed events is that the probability of the event happening in a short interval Δt is proportional to Δt. By the principle of mass action, this probability is also proportional to the number of molecules of the reactants. It follows that we can write:

$$
\begin{aligned}
&P(\text{a binding reaction happened in time } \Delta t \mid n_A + 1, \\
&\quad n_B + 1, n_C - 1 \text{ molecules at time } t) \\
&= \frac{k_1}{V}(n_A + 1)(n_B + 1)\Delta t,
\end{aligned} \tag{7}
$$

where V is the reaction volume.

Why do we write $\frac{k_1}{V}$ instead of just k_1? This is quite a subtle point of statistical physics (Van Kampen 1992), but it can be intuitively justified by a dimensional argument. Concentrations have units of *inverse volume*, so in order for Eq. (3) to be dimensionally consistent, k_1 must have units of *volume per time*, whereas k_2 simply has units of *inverse time*. The extra V is thus required to make Eq. (7) dimensionally consistent. It can generally be shown that reactions with n reactants will have their rates scaled by V^{1-n} (Van Kampen 1992). This can be intuitively justified by observing that bimolecular reactions will tend to happen less frequently in larger volumes, since it is harder for particles to find each other.

Applying the same arguments to scenarios (II) and (III), we find that we can write Eq. (4) as:

$$
\begin{aligned}
P(n_A, n_B, n_C; t + \Delta t) &= \frac{k_1}{V}(n_A + 1)(n_B + 1)\Delta t \, P(n_A + 1, n_B + 1, n_C - 1; t) \\
&+ k_2(n_C + 1)\Delta t \, P(n_A - 1, n_B - 1, n_C + 1; t) \\
&+ \left(1 - \frac{k_1}{V}n_A n_B \Delta t - k_2 n_C \Delta t\right) P(n_A, n_B, n_C; t). \tag{8}
\end{aligned}
$$

This equation simplifies very nicely to give:

$$
\begin{aligned}
&\frac{P(n_A, n_B, n_C; t + \Delta t) - P(n_A, n_B, n_C; t)}{\Delta t} \\
&= \frac{k_1}{V} \left[(n_A + 1)(n_B + 1)P(n_A + 1, n_B + 1, \right. \\
&\quad \left. n_C - 1; t) - n_A n_B P(n_A, n_B, n_C; t)\right] \\
&+ k_2 \left[(n_C + 1)P(n_A - 1, n_B - 1, \right. \\
&\quad \left. n_C + 1; t) - n_C P(n_A, n_B, n_C; t)\right]. \tag{9}
\end{aligned}
$$

The left-hand side of Eq. (9), in the limit of small Δt, is the definition of a derivative, and the right-hand side has no Δt dependence, so we get:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(n_A, n_B, n_C; t) = \frac{k_1}{V} \left[(n_A + 1)(n_B + 1) P(n_A + 1, n_B + 1, \right.$$
$$n_C - 1; t) - n_A n_B P(n_A, n_B, n_C; t)]$$
$$+ k_2 \left[(n_C + 1) P(n_A - 1, n_B - 1, n_C + 1; t) \right.$$
$$\left. -n_C P(n_A, n_B, n_C; t) \right], \tag{10}$$

which is known as the chemical master equation (CME) (Van Kampen 1992). Analogously to the REs, the CME can be solved numerically (Munsky and Khammash 2006), given an initial probability mass function $P(n_A, n_B, n_C; 0)$, or it can be solved analytically if it is sufficiently simple (Darvey et al. 1966; Gadgil et al. 2005; Jahnke and Huisinga 2007; Shahrezaei and Swain 2008; Grima et al. 2012; Smith and Shahrezaei 2015), which this example is (Darvey et al. 1966; Van Kampen 1976; Cianci et al. 2016). A less intensive approach is to approximate either the distribution (Thomas and Grima 2015; Smith et al. 2015; Andreychenko et al. 2017; Smith and Grima 2017b), or the moments (e.g. the mean or variance) of $P(n_A, n_B, n_C; t)$, and a variety of methods are popular such as the van Kampen approximation (Van Kampen 1992; Elf and Ehrenberg 2003), moment-closure approximations (Gillespie 2009b; Singh and Hespanha 2011; Grima 2012; Schnoerr et al. 2014a, 2015), or the chemical Langevin equation (Gillespie 2000; Schnoerr et al. 2014b). For a recent review see Ref. Schnoerr et al. (2017).

There is an alternative (and much more popular) approach to the CME, which has no analogue corresponding to the REs. This approach notes that $P(n_A, n_B, n_C; t)$ is not just any function, but is a probability distribution, and therefore, pseudo-random samples can be drawn from it exactly as one might sample from a Gaussian or a Poisson distribution (LEcuyer 2012). In fact, $P(n_A, n_B, n_C; t)$ is actually an infinite set of related probability distributions indexed by t, and so a sample will be an entire trajectory of molecule numbers over time. Such trajectories are much more intuitive than the CME itself, because each trajectory represents a particular realisation of what we might actually see if we observed a system in real time. Furthermore, it turns out it is typically computationally much easier to sample from the CME than to solve the CME, and if we take a large number of sample trajectories we can use them to estimate $P(n_A, n_B, n_C; t)$. The reason for this computational difference is that the cost of numerical solution of the CME scales with the number of possible system states (i.e. the number of permissible combinations of molecule numbers), while the cost of sampling scales with the frequency of reactions. Numerically solving the CME will typically only be worthwhile for systems with a small number of states and fast reaction rates.

Sampling trajectories of the CME is typically referred to as "simulating" the underlying system, for obvious reasons. The most common way to simulate is to use Gillespie's stochastic simulation algorithm (SSA) (Gillespie 1977), because this algorithm gives statistically exact trajectories, by taking advantage of the fact that the waiting times between reactions are exponentially distributed. The SSA can be quite slow, because it explicitly simulates every reaction, so a huge number of approximate simulation algorithms have been developed. The most popular is a time-discretised algorithm called τ-leaping, also by Gillespie (2001), but new algorithms are proposed
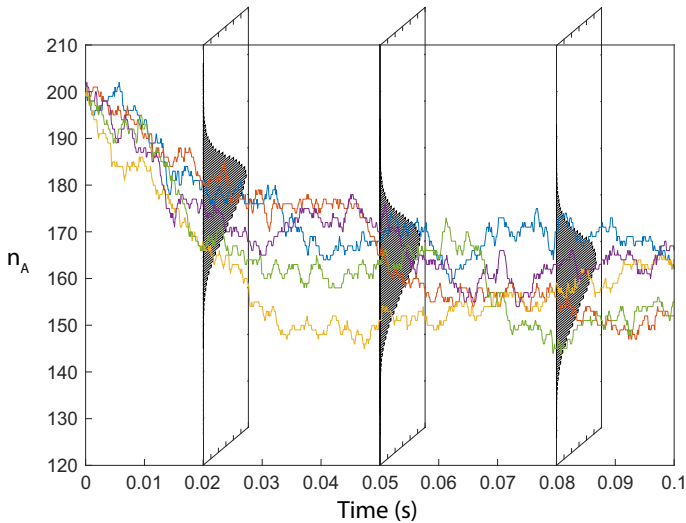
Fig. 2 Stochastic simulations and analytical solutions of Eq. (10). The five multicoloured trajectories are independent simulations made using the SSA. Exact probability distributions solving Eq. (10) are shown as perpendicular grey histograms at $t = 0.02$, 0.05 and 0.08. This makes clear that at each time point t the numerical value of each trajectory is a sample from the probability distribution solving the CME

every year that typically sacrifice a degree of accuracy for an increase in speed (Salis and Kaznessis 2005; Cao et al. 2005; Auger et al. 2006). For a recent review see Ref. Szekely and Burrage (2014).

The relationship between individual sample trajectories and the CME solution $P(n_A, n_B, n_C; t)$ is quite unintuitive, so we have shown an example in Fig. 2. The main graph shows five independent sample trajectories of the CME Eq. (10), showing the number of A molecules n_A over a short time period. Perpendicular to the main graph are three histograms showing the marginal probability distribution $P(n_A; t)$ at three time points $t = 0.02$, 0.05 and 0.08. This representation makes it clear that the values of the trajectories of n_A at $t = 0.02$ (for example) are independent samples of the distribution $P(n_A; 0.02)$. There is a one-to-one correspondence between the distributions $P(n_A, n_B, n_C; t)$ and the sample trajectories: if we know the distributions then we can sample trajectories; if we have enough independent trajectories we can approximate the distributions to an arbitrarily high degree of accuracy.

One point that we have not satisfactorily addressed so far is the validity of the principle of mass action, the principle which underpins fundamental equations such as (7). The principle of mass action is based on combinatorial arguments: if a reaction has two reactants, A and B say, then the probability of a reaction must be proportional to the number of ways a reacting pair can be made up, i.e. $n_A n_B$ different ways; similarly, if a reaction involves two molecules of A reacting with each other, then we would expect the reaction probability to scale as $\frac{n_A(n_A-1)}{2}$.

These arguments are not particularly satisfactory when considered from a microscopic point of view: if no pair of molecules is sufficiently close together to collide and react in a short time Δt, then a reaction will not happen, no matter how many
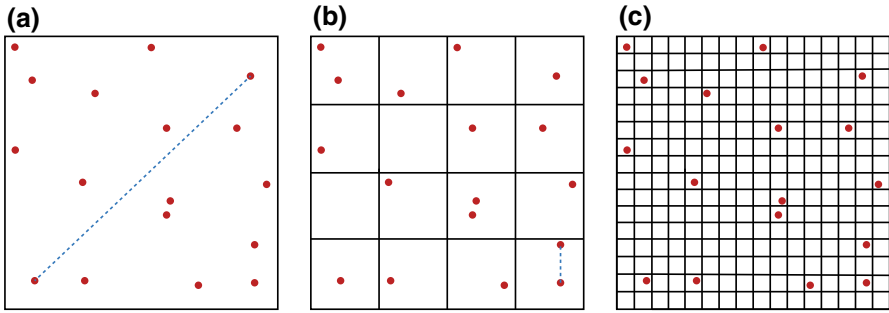
(a) **(b)** **(c)**

Fig. 3 The rationale behind dividing a volume into subvolumes. **a** The principle of mass action claims that every reacting pair of molecules is equally likely to react in any time period, including pairs which might be very distant (blue dotted line). **b** One solution is to divide the volume into M subvolumes, and only allow molecules in the same subvolume to react, so reacting pairs are now close together (blue dotted line). **c** If M is too large, then there will very rarely be pairs in the same subvolume, and bimolecular reactions will tend not to happen at all

molecules there are overall; or if only 2 (or 3 or 4 or …) pairs are sufficiently close together, then the probability of reaction will be roughly proportional to 2 (or 3 or 4 or …) rather than the total number of pairs. The principle of mass action surely amounts to assuming (nonsensically) that *all* the pairs are sufficiently close together to potentially react in a short time Δt. The absurdity of this assumption is demonstrated graphically in Fig. 3a. A counter-argument to this is that we have no way of knowing how many pairs are close together, so we stick to assuming that a fixed proportion of them are close together and absorb the proportionality constant into the reaction rate. But this again is not truly satisfactory because the true proportion of pairs that are close together is itself a random variable and not a constant.

It turns out that the principle of mass action is closely related to the *diffusion* of the reacting molecules. The diffusion coefficient is essentially a measure of the rate at which a randomly diffusing particle moves around the reaction volume. A particle with a very high diffusion coefficient could have diffused all over the volume in a short time Δt, whereas a particle with a small diffusion coefficient may only have covered a small region of space. (The average distance covered by a diffusing molecule before it reacts is known as the *Kuramoto length*, see Ref. Grima and Schnell 2008.) It follows that the principle of mass action could plausibly apply if particles have very high diffusion coefficients, so that any pair could potentially collide in a short time Δt. This idea can actually be proved from a microscopic point of view (and we will demonstrate this later), but *only if* we assume all particles have infinite (or practically infinite) diffusion coefficients. (Of course, the concept of an "infinite diffusion coefficient" is physically impossible, and we will address this point shortly.) Doubting the principle of mass action is then essentially equivalent to believing that the diffusion coefficients might not be high enough for the reaction volume to be well mixed, and this is a sensible belief as far as cell biology is concerned (Wojcieszyn et al. 1982). Clearly, an alternative approach is needed: in particular, we want a model which can tell us how many pairs of particles are "close together" at any given time.

The simplest way to go about this is to subdivide the entire reaction volume into small subvolumes, with the implication that we will consider molecules in the same

subvolume to be "close together", while molecules in different subvolumes are not. Suppose we have chosen to divide our volume into M subvolumes, which for simplicity we assume are equally sized; then, the probability of a reaction involving A and B in subvolume i will be proportional to $n_A^{(i)} n_B^{(i)}$, where $n_X^{(i)}$ denotes the number of molecules of species X in subvolume i. This idea is shown in Fig. 3b.

While superficially satisfying, this description actually raises more questions than it answers, at least at first. Principally, how should we choose M? Clearly, we should choose $M > 1$; otherwise, the description is identical to the CME, and generally, if M is small we will tend to have the same issues with the principle of mass action that the CME had, which drove us to seek a new model. On the other hand, if we make M very large, the subvolumes will become very small (possibly smaller than the physical size of a molecule) and the probability of two reacting molecules being in the same subvolume at a given time becomes negligible. As shown in Fig. 3c, this results in bimolecular reactions simply not happening, which is clearly something we would like to avoid.

Choosing the correct partition of the volume into subvolumes requires us to be more specific about what we truly mean by "close together", a phrase about which we have thus far been deliberately vague. To do this, we would need to propose a physical model for how chemical reactions occur, including molecular sizes and shapes, how they diffuse, whether they react immediately upon collision or whether they need a sufficiently large kinetic energy, and how electrostatic and hydrodynamic interactions impact on the reaction. These are questions that go far beyond the scope of our models, which would significantly detract from their simplicity. As a result, the standard response is to be vague about the value of M, suggesting that it should be neither too small nor too large. Though this is not a very satisfying answer, it highlights that there is no value of M which can be pre-specified for all situations: the correct M will depend on a number of factors and will likely be different for different situations, such as different diffusion coefficients.

Another significant issue with this model is how we decide how many molecules are in the ith subvolume. This is an issue because molecules, in reality, do not remain in the same location forever, but diffuse throughout the reaction volume—indeed, it was the issue of diffusion which led us to seek out a new model. To address this issue, drawing on the argument in Ref. Gillespie et al. (2014), we will consider a simplified one-dimensional volume and a single diffusing molecule currently located at a point x. It is well known that the probability density functions of diffusing point-particles obey a partial differential equation (PDE) called the diffusion equation:

$$\frac{\partial}{\partial t} p(x, t) = D \frac{\partial^2}{\partial x^2} p(x, t), \qquad (11)$$

where $p(x, t)$ is the probability density of finding the particle at location x at time t, and D is the particle's diffusion coefficient. We can rewrite Eq. (11) in terms of small increments Δt and Δx and rearrange to get:

$$p(x, t + \Delta t) = p(x, t) + \frac{D \Delta t}{\Delta x^2} [p(x + \Delta x, t) - 2p(x, t) + p(x - \Delta x, t)]. \quad (12)$$

Now, consider a partition of the total volume into M subvolumes of incremental width Δx, and let subvolume k be centred around the point x. If we denote by $q(k, t)$ the probability that the molecule is in subvolume k at time t, then we will have that:
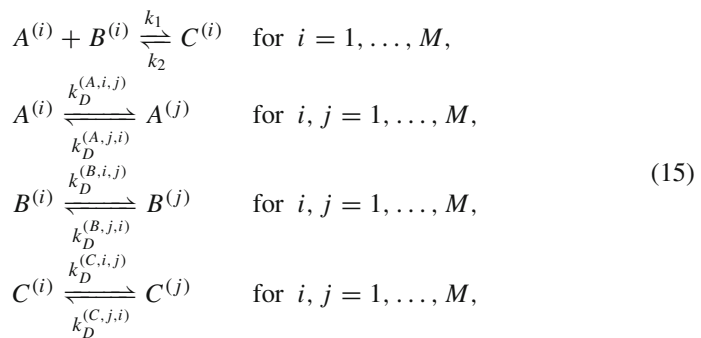
$$q(k, t) = \int_{x-\Delta x/2}^{x+\Delta x/2} p(y, t) \mathrm{d}y \approx p(x, t) \Delta x, \tag{13}$$

where the approximation will tend to hold when Δx is small. It follows that we can write:

$$q(k, t + \Delta t) = q(k, t) + \frac{D \Delta t}{\Delta x^2} \left[q(k+1, t) - 2q(k, t) + q(k-1, t) \right]. \tag{14}$$

In other words, in a short time Δt, the probability that the molecule moves from subvolume k to subvolume $k + 1$ is equal to $\frac{D \Delta t}{\Delta x^2}$. Equation (14) has the form of a master equation, which implies that the waiting time for a molecule diffusing between subvolumes is approximately exponential, at least when the subvolumes are small, but it is a reasonable approximation even when the subvolumes are quite large.

This leads us to a model in which particles "hop" between neighbouring subvolumes at random times, and the waiting times between hoppings are exponentially distributed. This means that we can represent hopping events as just another type of reaction event. This seems a little counter-intuitive, but makes sense when studied in detail. Let us consider system (2) again under this model. We can write $X^{(i)}$ to denote the species X in subvolume i. Then we can write the new system in the following way:

$$
\begin{aligned}
A^{(i)} + B^{(i)} &\underset{k_2}{\overset{k_1}{\rightleftharpoons}} C^{(i)} &&\text{for } i = 1, \ldots, M, \\
A^{(i)} &\underset{k_D^{(A,j,i)}}{\overset{k_D^{(A,i,j)}}{\rightleftharpoons}} A^{(j)} &&\text{for } i, j = 1, \ldots, M, \\
B^{(i)} &\underset{k_D^{(B,j,i)}}{\overset{k_D^{(B,i,j)}}{\rightleftharpoons}} B^{(j)} &&\text{for } i, j = 1, \ldots, M, \\
C^{(i)} &\underset{k_D^{(C,j,i)}}{\overset{k_D^{(C,i,j)}}{\rightleftharpoons}} C^{(j)} &&\text{for } i, j = 1, \ldots, M,
\end{aligned}
\tag{15}
$$

where $k_D^{(X,i,j)}$ is the "hopping rate" at which a molecule of type X in subvolume i will hop into subvolume j, which will be equal to zero if i and j are not neighbouring subvolumes. Note that we do not use the specific rate obtained in Eq. (14), because that rate is derived assuming that the subvolumes are arranged in a particular regular manner, and we would like to retain generality for the time being. Clearly, system (15) is much more complicated than (2), but there is little actual difference in principle. Both systems comprise species, reactions and rates—it just happens that the "species" in system (15) are not just the types of molecules (A, B, C), but the types of molecules in a particular subvolume ($A^{(i)}$, $B^{(i)}$, $C^{(i)}$). This might be unintuitive, but it makes no

difference mathematically. It follows, then, that any technique which can be applied to system (2) can also be applied to system (15).

For example, analogously to the REs (3), we can write equations for the concentrations $[A^{(i)}]$, $[B^{(i)}]$ and $[C^{(i)}]$ (intuitively, the local concentrations of A, B and C in subvolume i):

$$\frac{d[A^{(i)}]}{dt} = -k_1[A^{(i)}][B^{(i)}] + k_2[C^{(i)}] + \sum_{j=1}^{M}\left[-k_D^{(A,i,j)}[A^{(i)}] + k_D^{(A,j,i)}[A^{(j)}]\right],$$

$$\frac{d[B^{(i)}]}{dt} = -k_1[A^{(i)}][B^{(i)}] + k_2[C^{(i)}] + \sum_{j=1}^{M}\left[-k_D^{(B,i,j)}[B^{(i)}] + k_D^{(B,j,i)}[B^{(j)}]\right],$$

$$\frac{d[C^{(i)}]}{dt} = k_1[A^{(i)}][B^{(i)}] - k_2[C^{(i)}] + \sum_{j=1}^{M}\left[-k_D^{(C,i,j)}[C^{(i)}] + k_D^{(C,j,i)}[C^{(j)}]\right], \quad (16)$$

where, as before, M is the number of subvolumes. Equation (16) is the analogue of the REs with spatial resolution, but it would be helpful if we had a better idea of the form of the hopping rates $k_D^{(X,i,j)}$. Calculating the hopping rates from microscopic principles is an extremely challenging problem and, in fact, is still an open research question. In the simplest case, where each subvolume is an equally sized cube (or square, or line segment) of side length h arranged in a Cartesian grid, then we have already seen that there is a simple expression:

$$k_D^{(X,i,j)} = \begin{cases} \frac{D_X}{h^2} & \text{if } i \text{ neighbours } j, \\ 0 & \text{otherwise,} \end{cases} \quad (17)$$

where D_X is the diffusion coefficient associated with particles of type X.

If we imagine a one-dimensional array of subvolumes of equal size arranged in a line of length L, such that subvolume 1 neighbours subvolume 2, subvolume 2 neighbours subvolumes 1 and 3, subvolume 3 neighbours 2 and 4, etc., then Eq. (16) becomes:

$$\frac{d[A^{(i)}]}{dt} = -k_1[A^{(i)}][B^{(i)}] + k_2[C^{(i)}] + \frac{D_A}{(L/M)^2}\left[[A^{(i-1)}] - 2[A^{(i)}] + [A^{(i+1)}]\right],$$

$$\frac{d[B^{(i)}]}{dt} = -k_1[A^{(i)}][B^{(i)}] + k_2[C^{(i)}] + \frac{D_B}{(L/M)^2}\left[[B^{(i-1)}] - 2[B^{(i)}] + [B^{(i+1)}]\right],$$

$$\frac{d[C^{(i)}]}{dt} = k_1[A^{(i)}][B^{(i)}] - k_2[C^{(i)}] + \frac{D_C}{(L/M)^2}\left[[C^{(i-1)}] - 2[C^{(i)}] + [C^{(i+1)}]\right],$$

$$(18)$$

for $i = 2, \ldots, M - 1$ with small modifications for the end subvolumes 1 and M depending on the boundary conditions. We can now imagine taking the limit $M \to \infty$. There are a couple of issues with this. First, we already noted that choosing subvolumes too small could be a problem because the probability of two molecules being in the same subvolume would become negligibly small, thus making bimolecular reactions

unusually rare events. We will bypass this issue by saying that the limit $M \to \infty$ is simply an approximation, which we expect to be accurate when concentrations are high. The second issue is that the notation $[A^{(i)}]$ becomes meaningless when M is infinite. This we will solve by replacing $[A^{(i)}]$ with $[A]$ which we consider to be a function of location $x = \frac{iL}{M}$ as well as time. In the limit $M \to \infty$, x approaches a continuous quantity, so that differentiation with respect to x becomes valid. For example, $\frac{D_A}{(L/M)^2}\left[[A^{(i-1)}] - 2[A^{(i)}] + [A^{(i+1)}]\right]$ converges to $D_A \frac{\partial^2 [A]}{\partial x^2}$ in the limit $M \to \infty$. As a result, we get the following PDEs:

$$\frac{\partial [A]}{\partial t} = -k_1[A][B] + k_2[C] + D_A \frac{\partial^2 [A]}{\partial x^2},$$
$$\frac{\partial [B]}{\partial t} = -k_1[A][B] + k_2[C] + D_B \frac{\partial^2 [B]}{\partial x^2},$$
$$\frac{\partial [C]}{\partial t} = k_1[A][B] - k_2[C] + D_C \frac{\partial^2 [C]}{\partial x^2}. \tag{19}$$

These PDEs are the well-known reaction–diffusion equations (RDEs), ubiquitous in mathematical biology (Murray 2001), and popularised by Alan Turing in his seminal paper Ref. Turing (1952). Turing famously demonstrated that certain systems (Gierer and Meinhardt 1972) could be unstable when modelled with the RDEs, but stable when modelled with the REs: this kind of instability manifests itself as visual patterns (e.g. spots, stripes) in an RDE simulation. The Turing instability (as it is now known) is believed by some to be the cause of biological patterns such as zebrafish stripes (Nakamasu et al. 2009) or the regular spacing between mammalian digits (fingers and toes) (Sheth et al. 2012; Raspopovic et al. 2014), but there is still controversy around whether the Turing mechanism is really behind these phenomena (Watanabe and Kondo 2015). One of the current biggest challenges in synthetic biology is therefore to synthesise a Turing patterning network in living cells (Lengyel and Epstein 1992; Diambra et al. 2014; Borek et al. 2016; Scholes and Isalan 2017; Smith and Dalchau 2018a, b), which would provide convincing evidence of Turing's theory.

As well as the REs, we can also apply the CME methodology to system (15). Instead of the probability mass function $P(n_A, n_B, n_C; t)$, we now consider a new probability mass function $P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t)$, where $\vec{n}_X = (n_X^{(1)}, \ldots, n_X^{(M)})$ is a vector of molecule numbers, with one entry for each subvolume. Because of the complexity of the RDME, we have to introduce some new notations to be able to write it down compactly: we let $E_{(X,i)}$ be the shift operator which replaces any instance of $n_X^{(i)}$ with $n_X^{(i)} + 1$. For example: $E_{(X,i)} n_X^{(i)} = n_X^{(i)} + 1$, and, $E_{(X,i)}^{-1} n_X^{(i)} = n_X^{(i)} - 1$, and for any function $f(\cdot)$, $E_{(X,i)} f(n_X^{(i)}) = f(n_X^{(i)} + 1)$. Note that the inverse operator $E_{(X,i)}^{-1}$ replaces $n_X^{(i)}$ with $n_X^{(i)} - 1$.

Following the argument for the CME, we consider a time-step Δt, sufficiently short that at most one reaction (including hopping events) can occur. Then, we consider the ways we can end up with $\vec{n}_A, \vec{n}_B, \vec{n}_C$ molecules at time $t + \Delta t$. This can happen if: (I) a binding reaction happens in subvolume i, for some $i = 1, \ldots, M$, in the time interval $[t, t + \Delta t)$; (II) an unbinding reaction happens in subvolume i, for some $i = 1, \ldots, M$,

in $[t, t + \Delta t]$; (III) a particle of type X hops from subvolume i to subvolume j, for some $i, j = 1, \ldots, M$ and $X \in \{A, B, C\}$, in $[t, t + \Delta t]$; or (IV) no reactions happen in the time interval $[t, t + \Delta t]$. Clearly, this is much more complicated than for the CME, but exactly the same principles apply, and after some simplifications we obtain the following equation:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t) = {} & \sum_{i=1}^{M} \frac{k_1 M}{V} \left[E_{(A,i)} E_{(B,i)} E_{(C,i)}^{-1} - 1 \right] n_A^{(i)} n_B^{(i)} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t) \\
& + \sum_{i=1}^{M} k_2 \left[E_{(A,i)}^{-1} E_{(B,i)}^{-1} E_{(C,i)} - 1 \right] n_C^{(i)} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t) \\
& + \sum_{i=1}^{M} \sum_{j=1}^{M} k_D^{(A,i,j)} \left[E_{(A,i)} E_{(A,j)}^{-1} - 1 \right] n_A^{(i)} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t) \\
& + \sum_{i=1}^{M} \sum_{j=1}^{M} k_D^{(B,i,j)} \left[E_{(B,i)} E_{(B,j)}^{-1} - 1 \right] n_B^{(i)} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t) \\
& + \sum_{i=1}^{M} \sum_{j=1}^{M} k_D^{(C,i,j)} \left[E_{(C,i)} E_{(C,j)}^{-1} - 1 \right] n_C^{(i)} P(\vec{n}_A, \vec{n}_B, \vec{n}_C; t),
\end{aligned}
$$

$$(20)$$

noting the correct volume scaling $\frac{k_1 M}{V}$ for the bimolecular reaction. This equation is known as the reaction–diffusion master equation (RDME). The first two lines correspond to the two reactions in each subvolume, while the final three lines correspond to the hopping of particles of type A, B and C, respectively, between neighbouring subvolumes.

Now that we have the RDME (which is really just a special type of CME); we can do to it anything that we could do to the CME. For instance, it is possible to solve RDMEs analytically, if they are composed exclusively of certain types of linear reactions (Jahnke and Huisinga 2007; Gadgil et al. 2005) or purely reversible reactions (Cianci et al. 2016). The example (15) is actually solvable, as long as the hopping events are reversible (i.e. as long as $k_D^{(X,i,j)} = k_D^{(X,j,i)}$, for all $X = A, B, C$ and $i, j = 1, \ldots, M$). Other analytical approximation techniques for the CME which are beyond the scope of this article [e.g. the van Kampen approximation (1992) and the chemical Langevin equation (Gillespie 2000)] can also be applied to the RDME, and have been, with interesting results (Smith et al. 2016; Ghosh et al. 2015).

Any stochastic simulation algorithm designed for the CME can also naturally be applied to the RDME (Stundzia and Lumsden 1996; Elf and Ehrenberg 2004; Bernstein 2005; Fange et al. 2012). Several new issues arise when simulating the RDME, however. For example, stochastic simulation algorithms tend to have computation time proportional to the frequency of reaction events (Gillespie 1977). If the hopping rates are quite large, then an RDME simulation may take substantially longer than an equivalent CME simulation—it is not uncommon for upwards of 99% of the simulation time
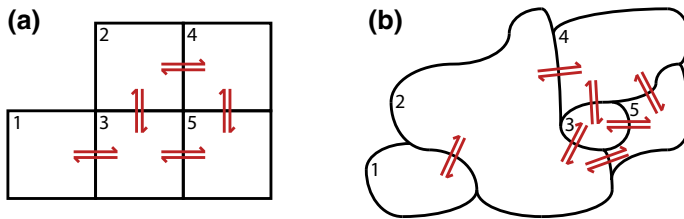
Fig. 4 **a** If subvolumes are arranged in a grid, all hopping rates are straightforwardly given by Eq. (17). **b** If subvolumes are of irregular shapes, and boundaries between subvolumes are of different sizes, it is not clear how to best choose hopping rates

to be spent simulating hopping events rather than chemical reactions. This issue in particular has driven a huge amount of research into fast and accurate simulation of the RDME (Elf et al. 2003; Drawert et al. 2010; Roberts et al. 2013; Fu et al. 2014).

It is worth going into a little more detail now about the hopping rates $k_D^{(X,i,j)}$. In the simplest case where each subvolume is of the same size, Eq. (17) can be used. This case is demonstrated in Fig. 4a: neighbouring subvolumes (such as 1 and 3) all have the same hopping rates between them, whereas non-neighbouring subvolumes (such as 1 and 2) have zero hopping rates. More generally, subvolumes can be of any shape and size, and the size of the boundaries where two neighbouring subvolumes join can vary greatly as well. Such a case is shown in Fig. 4b: neighbouring subvolumes (such as 1 and 2) tend to be of different sizes, so the hopping rate in one direction will generally be different from the other; moreover, boundaries between neighbours are not necessarily of the same length (such as the boundaries between 2 and 4, and 2 and 5), and shorter boundaries will tend to be crossed by fewer molecules, and so should have correspondingly lower hopping rates. In these cases, the hopping rate should (in principle) be calculated using a first passage time approach with the diffusion equation (Redner 2001), but this is often computationally unfeasible and in practice there is no easy answer. Some authors have proposed techniques for choosing these rates in a manner which agrees optimally with Brownian diffusion (Isaacson and Peskin 2006; Engblom et al. 2009; Bayati et al. 2011; Drawert et al. 2012; Lotstedt and Meinecke 2015; Meinecke et al. 2016; Meinecke and Ltstedt 2016), but these are typically simulation-based and not analytically straightforward.

In this section, we typically assume that molecules do not occupy any volume themselves (i.e. they are point-particles), but relaxing this assumption leads to even more complex hopping rates. These rates typically depend on the concentrations of the various species in the destination subvolume j, so that a subvolume with a high concentration of molecules is less likely to be able to accommodate a new molecule hopping into it (Elderfield 1985; Baker et al. 2010; Fanelli and McKane 2010; Cianci et al. 2016). Writing the hopping rate as a linear function of the concentrations is a popular choice, but when concentrations are very high certain nonlinear functions have been shown to be more accurate (Meinecke 2017; Cianci et al. 2017; Smith et al. 2017).

When diffusion coefficients are large, simulations of the RDME can become very slow (because molecules will hop around many times between reactions) (Drawert

et al. 2010), and significant modifications to the hopping rates have been introduced to address this issue. In particular, it is common to allow particles to hop between non-neighbouring subvolumes i and j (Taylor et al. 2014). The implication is that the particle did in fact follow a neighbour-to-neighbour path from i to j, but the intermediate steps are eliminated for speed: a very careful choice of the hopping rates is required in these cases to correctly model diffusion.

We have seen how the RDME was obtained as an extension of the CME, to address the problem of mass action kinetics. In this sense, the RDME is a *top-down* model, since it is not a simplification of a more detailed microscopic model like Brownian dynamics. The risk with top-down models is that the new components (the hopping rates, in the case of the RDME) could be added in a flawed manner, making the model inconsistent with the other well-established models. In short, the RDME is just the CME with some additional terms to explicitly model diffusion, but how do we know that these additional terms are correct?

This question has plagued users of the RDME for decades. A whole subgenre of the reaction–diffusion field is dedicated to comparing the RDME with microscopic models, including various types of Brownian dynamics and molecular dynamics, which are taken to be the "ground truth". Almost invariably, because of the complexities of the models involved, it is stochastic simulations of models which are compared, rather than the actual equations which make up the models (Baras and Mansour 1996; Dobrzynski et al. 2007; Fange et al. 2010). As a result, the outcome of these comparisons is typically to suggest optimal values of the various RDME model parameters (reaction rates, hopping rates, number and size of subvolumes) to get best agreement with the "ground truth" model, for a particular reaction–diffusion system. Such comparisons are therefore very limited in their applicability: there is no reason to believe that the optimal hopping rates (for example) for one system will be the same for another system, even if they are quite similar.

An alternative (though much trickier) approach is to directly compare the RDME with the "ground truth" model by writing out equations for both and discerning to what extent they agree (Collins and Kimball 1949; Fange et al. 2010; Hellander et al. 2015). This kind of comparison will provide general rules about when models agree and will not generally be limited by system specifics. The challenge is that "ground truth" models tend to be very hard to write down, since they are very complex and are usually conceived as models for simulation rather than mathematical analysis. A pioneer in this kind of comparison is Isaacson, who has published several articles analytically comparing the RDME with a variety of "ground truth" models (Isaacson 2008, 2009, 2013). Two principal results emerge from Isaacson's work. First, the RDME can be thought of as an approximation to a particular BD model known as the Smoluchowski model (1917) (more on this later), but there is no limit in which one converges to the other (Isaacson 2009). It can be rigorously proved that as M increases, the approximation first gets better, and then gets worse (while never actually showing perfect agreement), and it is hard to know in general which M is the optimal one. In a similar vein, Hellander et. al. have shown that there exists a hard upper limit on M, above which the RDME cannot agree with the Smoluchowski model because the subvolumes become unphysically smaller than the size of a molecule (Hellander et al. 2012). Isaacson's second principal result is that the RDME can be modified to create

a new model (known as the "convergent RDME", or CRDME) in which molecules can react with molecules in different subvolumes, according to certain strict rules (Isaacson 2013). The CRDME was shown to converge precisely to another BD model known as the $\lambda - \rho$ model (Erban and Chapman 2009) (more on this later). Overall, it seems that the RDME never quite sits comfortably with BD, and potentially, this could call its accuracy into question.

But we take a slightly different view of the RDME to those who wish to compare it to a "ground truth" model. Considering the RDME as an approximation to a microscopic model necessarily involves questions about molecular shapes and sizes and about the physical chemistry of reactions, questions which take us far from the neat simplicity of the CME. The RDME was conceived as a version of the CME with spatial resolution, to address the question of how diffusion affects chemical systems. Users of the RDME are not necessarily interested in (or even aware of) the size and shape of the molecules they are modelling, and to require knowledge of these quantities to correctly parameterise the RDME is a potential overcomplication. In our view, the RDME is a neat and straightforward model for *qualitatively* understanding the effect of diffusion on stochastic reaction systems, an important academic question for which no other model is well suited, but it is probably not appropriate to answer questions where numerical accuracy are paramount, and in these cases a microscopic model may be more suitable.

We do not worry particularly about the RDME's agreement with more complex models, to the extent that we do not discredit it as a model simply because it disagrees with more microscopic models, but in Sect. 4 we do try to analyse *why* the RDME differs from BD. However, it is very important that the RDME should agree with the simpler models on the complexity scale in the conditions under which those models are accurate. In particular, we would like the RDME to agree with the CME when diffusion is very fast (i.e. when spatial resolution becomes unimportant); we would like the RDME to agree with the RDEs when concentrations are very high (i.e. when stochasticity becomes unimportant); and we would like it to agree with the REs when both of the above conditions hold (high concentrations and fast diffusion).

2.1 The Limit of Fast Diffusion

We have previously mentioned the concept of "infinite diffusion", which is somewhat counter-intuitive and is worth discussing. Diffusion coefficients in reality cannot be infinite. In the case of spherical particles, they are given by the Stokes–Einstein relation $D = \frac{k_B T}{6\pi \eta r}$, where k_B is the Boltzmann constant, T is the temperature, η is the viscosity of the solvent, and r is the particle radius (Gillespie and Seitaridou 2012). All of these quantities are finite and nonzero, so infinite diffusion is necessarily unphysical. When we say "infinite diffusion" we are not thinking of limiting case of small particle radii, low viscosity or high temperatures. Instead, we mean that the rate at which a particle moves between regions of space (the hopping rate, in the RDME) is much higher than the rate at which it is involved in chemical reactions (these conditions are typically referred to as "reaction-limited" (Grima and Schnell 2006), but thinking instead in terms of "infinite diffusion" will prove to be useful shortly). In other words, a typical

particle will diffuse throughout the reaction volume several times before actually reacting. Under these conditions, the principle of mass action kinetics can reasonably be said to hold, since any pair of particles will have time to move close together before a reaction takes place, and the concept of particles being too far apart to react no longer makes sense. For this reason, we expect the CME to be an accurate model under the conditions of "infinite diffusion", and so if the RDME is also accurate, it should agree with the CME in this case.

To answer the question of agreement between the CME and RDME, we will study a straightforward but general example. For simplicity, we will consider a system consisting of only one chemical species, A, and two equally sized subvolumes inside a total volume of size V. The set-up is shown in Fig. 5a. The arguments that follow in fact apply to systems with any number of species and any number of subvolumes of any sizes, and a complete proof can be found in Ref. Smith and Grima (2016).

The system will consist of R reactions which take the form $s_j A \rightarrow r_j A$ for $j = 1, \ldots, R$. The rate of reaction j will be given by an arbitrary function $f_j(n_A, V)$. The CME of our system takes the form:

$$\frac{d}{dt} P(n_A; t) = \sum_{j=1}^{R} \left[f_j(n_A + s_j - r_j, V) P(n_A + s_j - r_j; t) - f_j(n_A, V) P(n_A; t) \right].$$
(21)

In contrast, the RDME takes the form:

$$\frac{d}{dt} P\left(n_A^{(1)}, n_A^{(2)}; t\right) = \sum_{j=1}^{R} \left[f_j\left(n_A^{(1)} + s_j - r_j, \frac{V}{2}\right) P\left(n_A^{(1)} + s_j - r_j, n_A^{(2)}; t\right) \right.$$
$$\left. - f_j\left(n_A^{(1)}, \frac{V}{2}\right) P\left(n_A^{(1)}, n_A^{(2)}; t\right) \right]$$
$$+ \sum_{j=1}^{R} \left[f_j\left(n_A^{(2)} + s_j - r_j, \frac{V}{2}\right) P\left(n_A^{(1)}, n_A^{(2)} + s_j - r_j; t\right) \right.$$
$$\left. - f_j\left(n_A^{(2)}, \frac{V}{2}\right) P\left(n_A^{(1)}, n_A^{(2)}; t\right) \right]$$
$$+ k_D^{(A,1,2)} \left[\left(n_A^{(1)} + 1\right) P\left(n_A^{(1)} + 1, n_A^{(2)} - 1; t\right) \right.$$
$$\left. - n_A^{(1)} P\left(n_A^{(1)}, n_A^{(2)}; t\right) \right]$$
$$+ k_D^{(A,2,1)} \left[\left(n_A^{(2)} + 1\right) P\left(n_A^{(1)} - 1, n_A^{(2)} + 1; t\right) \right.$$
$$\left. - n_A^{(2)} P\left(n_A^{(1)}, n_A^{(2)}; t\right) \right].$$
(22)

We note that the second argument of the f_j in the RDME is $\frac{V}{2}$ because each subvolume has volume $\frac{V}{2}$. Because each of the subvolumes is equal in size, and since they share a boundary, we can safely assume $k_D^{(A,1,2)} = k_D^{(A,2,1)}$. If this were not the case, it would imply a net flow of particles from one volume to the other, which may be appropriate

for modelling some problems, but not the typical problem of diffusion which has no directional bias.

The limit of infinite diffusion implies that lines 3 and 4 of Eq. (22) will tend to overwhelmingly dominate, unless $P(n_A^{(1)}, n_A^{(2)}; t)$ has a form which makes lines 3 and 4 equal to zero. It follows that the system will converge to this form (the equilibrium distribution of lines 3 and 4) infinitely quickly, i.e. the following equation holds at all times:

$$0 = k_D^{(A,1,2)} \left[\left(n_A^{(1)} + 1 \right) P \left(n_A^{(1)} + 1, n_A^{(2)} - 1; t \right) - n_A^{(1)} P \left(n_A^{(1)}, n_A^{(2)}; t \right) \right]$$
$$+ k_D^{(A,2,1)} \left[\left(n_A^{(2)} + 1 \right) P(n_A^{(1)} - 1, n_A^{(2)} + 1; t) - n_A^{(2)} P \left(n_A^{(1)}, n_A^{(2)}; t \right) \right]. \quad (23)$$

Since $k_D^{(A,1,2)} = k_D^{(A,2,1)}$, this simplifies further to:

$$0 = \left(n_A^{(1)} + 1 \right) P \left(n_A^{(1)} + 1, n_A^{(2)} - 1; t \right) - n_A^{(1)} P \left(n_A^{(1)}, n_A^{(2)}; t \right) + \left(n_A^{(2)} + 1 \right)$$
$$P \left(n_A^{(1)} - 1, n_A^{(2)} + 1; t \right) - n_A^{(2)} P \left(n_A^{(1)}, n_A^{(2)}; t \right). \quad (24)$$

This equation looks complicated, but in fact it has a straightforward non-trivial solution:

$$P \left(n_A^{(1)}, n_A^{(2)}; t \right) = \begin{cases} \frac{n_A! 2^{-n_A}}{n_A^{(1)}! n_A^{(2)}!} & \text{if } n_A^{(1)} + n_A^{(2)} = n_A, \\ 0 & \text{otherwise,} \end{cases} \quad (25)$$

where n_A is, as before, the total number of molecules of A in the entire volume. The solution (25) is a Binomial $\left(n_A, \frac{1}{2} \right)$ distribution, which is intuitively not surprising. If diffusion is infinitely fast, at any given time, each molecule has a 50% chance of being in subvolume 1 and a 50% chance of being in subvolume 2. If there are n_A molecules overall, then the solution must be the one given in Eq. (25).

We now take a moment to think about the meaning of the functions f_j. Since waiting times between reactions are exponentially distributed, $f_j(n_A, V)\Delta t$ is the probability that reaction j happens in a short time period of length Δt, given that there are n_A molecules in the volume V. f_j is precisely this probability in the CME, but the analogous expression for this probability in the RDME is not immediately obvious, because the RDME concerns the local molecule numbers $n_A^{(1)}$ and $n_A^{(2)}$ and the total molecule number n_A is never explicitly modelled. Let us nonetheless denote it by $\bar{f}_j(n_A, V)$. It follows that the total molecule number in the RDME with infinitely fast diffusion satisfies the following equation:

$$\frac{d}{dt} P(n_A; t) = \sum_{j=1}^{R} \left[\bar{f}_j(n_A + s_j - r_j, V) P(n_A + s_j - r_j; t) - \bar{f}_j(n_A, V) P(n_A; t) \right],$$
$$(26)$$

which is exactly the CME (21), but with f_j replaced by $\bar{f}_j$. There is a subtle point here: we are saying that we can rewrite the RDME in terms of the total molecule numbers

n_A, rather than the local molecule numbers $n_A^{(i)}$. We can do this because diffusion is infinitely fast, and so $n_A^{(i)}$ becomes a random variable parameterised by n_A, and the values of $n_A^{(i)}$ at any two distinct time points are uncorrelated. If diffusion were not infinitely fast, we would not be able to rewrite the RDME in the form of Eq. (26) without losing some information.

It remains to evaluate $\bar{f}_j$: if it is equal to f_j then we have proved that the RDME converges to the CME in the limit of fast diffusion. $\bar{f}_j(n_A, V)\Delta t$ is the probability that reaction j happens in a time period Δt and is equal to the probability that it happens in subvolume 1 or in subvolume 2. Since these events are independent, we get:

$$\bar{f}_j(n_A, V)\Delta t = \sum_{n_A^{(1)}=0}^{n_A} \sum_{n_A^{(2)}=0}^{n_A} P\left(n_A^{(1)}, n_A^{(2)}; t\right) \left[f_j\left(n_A^{(1)}, \frac{V}{2}\right) \Delta t + f_j\left(n_A^{(2)}, \frac{V}{2}\right) \Delta t \right]$$

$$= 2 \sum_{n_A^{(1)}=0}^{n_A} \sum_{n_A^{(2)}=0}^{n_A} P\left(n_A^{(1)}, n_A^{(2)}; t\right) f_j\left(n_A^{(1)}, \frac{V}{2}\right) \Delta t$$

$$= 2\mathbb{E}\left[f_j\left(n_A^{(1)}, \frac{V}{2}\right) \Big| n_A \right] \Delta t, \tag{27}$$

where the conditional expectation, $\mathbb{E}\left[\star | n_A\right]$, denotes the expected value of $\star$ given that there are n_A molecules overall. In short, $\bar{f}_j$ is proportional to the expected value of f_j under the binomial distribution (25). We can try some example f_j's to see what happens. For instance, if we choose $f_j(n_A, V) = k$ for some constant k, then we get:

$$\bar{f}_j = 2\mathbb{E}\left[k | n_A\right] = 2k \neq k = f_j, \tag{28}$$

so f_j cannot simply be a constant. On the other hand, if we choose the correct volume scaling for a zero-order reaction propensity, $f_j(n_A, V) = kV$, then we get:

$$\bar{f}_j = 2\mathbb{E}\left[k \frac{V}{2} | n_A\right] = 2k\frac{V}{2} = kV = f_j, \tag{29}$$

so the RDME and CME are consistent. This shows the importance of the correct volume scaling for reaction rates. Some other standard rates are the monomolecular rate $f_j(n_A, V) = kn_A$, corresponding to a reaction of the form $A \xrightarrow{k}$, which gives:

$$\bar{f}_j = 2\mathbb{E}\left[kn_A^{(1)} | n_A\right] = 2k\frac{n_A}{2} = kn_A = f_j, \tag{30}$$

and the bimolecular rate $f_j(n_A, V) = \frac{k}{V}n_A(n_A - 1)$, corresponding to a reaction of the form $A + A \xrightarrow{k}$, which gives:

$$\bar{f}_j = 2\mathbb{E}\left[\frac{2k}{V}n_A^{(1)}(n_A^{(1)} - 1) | n_A\right] = \frac{4k}{V}\left(\frac{n_A}{4} + \frac{n_A^2}{4} - \frac{n_A}{2}\right) = \frac{k}{V}n_A(n_A - 1) = f_j. \tag{31}$$

These standard rates provide consistency between the RDME and CME, although the original physical reasoning behind these volume scalings have nothing to do with the RDME (Van Kampen 1992). But even rates which scale correctly with volume can have problems if they are unphysical in other ways. For instance, if we chose $f_j(n_A, V) = \frac{k}{V} n_A^2$, which has the correct volume scaling for a bimolecular reaction, we would get:

$$\bar{f}_j = 2\mathbb{E}\left[\frac{2k}{V}\left(n_A^{(1)}\right)^2 | n_A\right] = \frac{4k}{V}\left(\frac{n_A}{4} + \frac{n_A^2}{4}\right) = \frac{k}{V} n_A(n_A + 1) \neq \frac{k}{V} n_A^2 = f_j. \tag{32}$$

Our RDME consistency criterion appears to very carefully select for rate functions with a sound physical basis, so it is clearly necessary to be very careful with rates before using the RDME.

While the "bad" rate functions given above are clearly just mistakes which can be rectified, there are some other genuine rate functions in common usage known as "non-elementary rates" which are not so fortunate. These include Hill function rates and Michaelis–Menten rates (Lawson et al. 2015). The Michaelis–Menten rate function has the form $f_j(n_A, V) = \frac{kVn_A}{KV+n_A}$, which gives:

$$\bar{f}_j = 2\mathbb{E}\left[\frac{kVn_A^{(1)}}{KV + 2n_A^{(1)}} | n_A\right] = \frac{kVn_A \, _2F_1\left(\frac{KV}{2} + 1, 1 - n_A; \frac{KV}{2} + 2; -1\right)}{2^{n_A}\left(\frac{KV}{2} + 1\right)}$$
$$\neq \frac{kVn_A}{KV + n_A} = f_j, \tag{33}$$

where $_2F_1(a, b; c; d)$ is a hypergeometric function (Abramowitz and Stegun 1964). The RDME rate in this case is so different from the CME rate that it could give very wrong results if used. To check this, we simulated the following system:

$$\emptyset \xrightarrow{k_1} A, \; A \xrightarrow{MM} \emptyset, \tag{34}$$

where $\emptyset$ denotes that we are not interested in the species involved, and MM denotes that the reaction occurs with the Michaelis–Menten rate $\frac{k_2 Vn_A}{KV+n_A}$. System (34) was simulated using both the RDME with $M = 2$ and fast diffusion, and the CME. The results are shown in Fig. 5. For the chosen parameters (see caption) the distributions are remarkably different: the RDME distribution has mean of around 7 compared with 11 for the CME. It is clear that using the RDME with non-elementary rates is a fundamentally bad idea. For more details of this effect, see Ref. Smith and Grima (2016).

We set out to demonstrate that the RDME is consistent with the CME when diffusion is infinitely fast, and we proved that this is the case but only under certain conditions. It appears that there is a strict subset of rate functions which are compatible with the RDME, and if a rate function is used which is not from this subset then the RDME is no longer a consistent model. This does not cast doubt on the RDME necessarily, but it does imply that extra caution should be taken when choosing rates for the RDME.
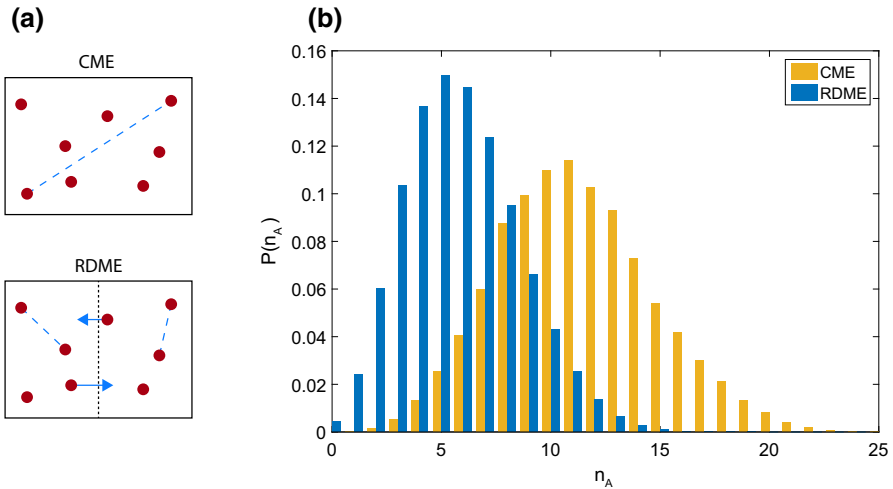
(a)

(b)

Fig. 5 Comparing the fast diffusion limit of the RDME with the CME. **a** In the CME, any molecule (red) can react with any other (dashed blue line); in the RDME molecules can only react with other molecules in the same subvolume (dashed blue lines), but they can "hop" between neighbouring subvolumes (blue arrows). **b** The equilibrium distribution of the CME (yellow) and RDME with $M = 2$ and fast diffusion ($k_D = 40$, blue) for system (34) obtained from stochastic simulations, with $k_1 = 0.1$, $k_2 = 1$, $K = 100$ and $V = 1$. If the RDME is correct these two distributions should essentially agree, but they are clearly very different. Note that the RDME distribution remains the same for larger k_D

2.2 The Limit of High Concentrations

As we have mentioned, the RDME is just a type of CME, and so any argument that holds for the CME will also hold for the RDME. The most fundamental result in the theory of stochastic chemical kinetics is that the CME converges to the REs when concentrations are high, thus implying that stochasticity stops being important for high concentrations. We will not go into the proof of this here, but it is discussed in detail in Refs. Kurtz (1971, 1972), Van Kampen (1992), Gardiner (1986), Ball et al. (2006) and Hepp et al. (2015). Naturally, this means that the RDME also converges to the spatially discretised analogue of the REs, Eq. (16), when concentrations are high. We will briefly offer a justification of this here, although it is not a rigorous proof by any means.

We will think about the number of A molecules in subvolume i at a time $t + \Delta t$, $n_A^{(i)}(t + \Delta t)$, conditional on the molecule numbers at time t. $n_A^{(i)}$ is typically a random variable, but when concentrations are high it will converge to its mean by the law of large numbers, so we can essentially treat it as a fixed value for now. Two facts help us to evaluate $n_A^{(i)}(t + \Delta t)$. Firstly, because there are huge numbers of molecules of all types in all subvolumes, a huge number of reaction events of each type will happen in the time period $[t, t + \Delta t]$. This implies that we do not need to think about the probability that a reaction happens, but the number of times it happens, and by the law of large numbers this number will converge (in the limit of high concentrations) to its expected value. Secondly, even though a huge number of reactions will happen in

$[t, t + \Delta t)$, Δt is still small, so the total molecule numbers will not change significantly over that period. This means that we can use $n_A^{(i)}(t)$ to refer to the value of $n_A^{(i)}$ at any point in the interval $[t, t + \Delta t)$. Following this argument, for system (2) in a volume V with length L divided linearly into M subvolumes, thinking about the rates of the processes which can increase or decrease $n_A^{(i)}$, we find that we can write:

$$n_A^{(i)}(t + \Delta t) = n_A^{(i)}(t) - \frac{k_1 M}{V} n_A^{(i)}(t) n_B^{(i)}(t) \Delta t + k_2 n_C^{(i)}(t) \Delta t$$
$$+ \frac{D_A M^2}{L^2} \left[n_A^{(i-1)}(t) - 2n_A^{(i)}(t) + n_A^{(i+1)}(t) \right] \Delta t. \qquad (35)$$

Note that we have used the volume scaling $\frac{k_1 M}{V}$ for the bimolecular reaction. Using the same arguments as in the derivation of the CME, Eq. (4), we can rearrange the Δt's in Eq. (35) to produce a differential equation:

$$\frac{d}{dt} n_A^{(i)} = -\frac{k_1 M}{V} n_A^{(i)} n_B^{(i)} + k_2 n_C^{(i)} + \frac{D_A M^2}{L^2} \left[n_A^{(i-1)} - 2n_A^{(i)} + n_A^{(i+1)} \right]. \qquad (36)$$

We can divide through by the volume of the subvolumes $\frac{V}{M}$ to get the concentrations:

$$\frac{d}{dt} [A^{(i)}] = - k_1 [A^{(i)}][B^{(i)}] + k_2 [C^{(i)}] + \frac{D_A M^2}{L^2} \left[[A^{(i-1)}] - 2[A^{(i)}] + [A^{(i+1)}] \right]. \qquad (37)$$

Note that the volume scaling of the bimolecular rate disappears when we convert to concentrations. Equation (37) is exactly Eq. (18), which is the spatial analogue of the REs. We have already discussed that, in the limit of high concentrations, we can justifiably take the limit $M \to \infty$ to produce the RDEs Eq. (19), so the RDEs turn out to be the limiting form of the RDME when concentrations are high.

2.3 The Combined Limits of Fast Diffusion and High Concentrations

We have established that the RDME reduces exactly to the CME in the limit of fast diffusion (if the rates are physically valid) or the REs in the limit of high concentrations; in this section, we will briefly consider what happens when *both* of these limits are applied simultaneously.

We have already noted that the high concentration limit of the CME is the REs, and we have a rough sketch of why this is the case. We will not go into the subject again here, and the proof can be found in Refs. Van Kampen (1992) and Gardiner (1986)

But what happens to the RDEs when diffusion is fast? The RDEs have the form:

$$\frac{\partial}{\partial t} [A] = R([A]) + D_A \nabla^2 [A], \qquad (38)$$

where $R(\cdot)$ denotes reaction terms. When diffusion is fast, the second term of (38) dominates unless $\nabla^2 [A] = 0$, which happens if $[A]$ is a spatially uniform concentration. When diffusion is infinitely fast, the concentration $[A]$ will converge to a spatially

Fig. 6 How the RDME fits into the scale of model complexity shown in Fig. 1. There is no direct relationship between the RDME and any of the more complex models
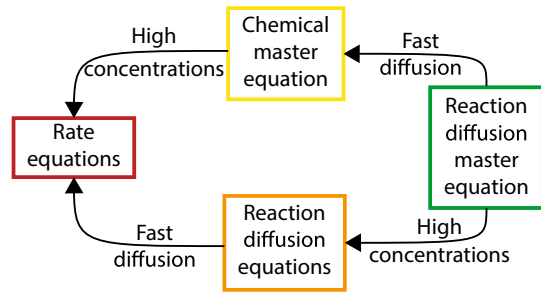
uniform concentration infinitely quickly. In that case, $[A]$ will have no spatial dependence and will simply become a function of t. The RDEs then obey:

$$\frac{\partial}{\partial t}[A] = R([A]),\tag{39}$$

which are simply the REs. It does not matter, therefore, in which order the limits of fast diffusion and high concentrations are applied: either way the RDME becomes the REs.

Overall, we have found that the RDME can be thought of as a "parent model" of the CME, the RDEs and the REs, with the relationship between them shown in Fig. 6. Specifically, if we use both the RDME and the CME (for example) to model the same system, and the system is such that we expect the CME to be correct (i.e. diffusion is fast), then the two models will give identical predictions. This is great for the RDME and demonstrates its place in the complexity scale (Fig. 1), but the RDME is nonetheless a top-down model with no guaranteed microphysical basis. In the next section, we investigate what happens if we try to reach the same point on the complexity scale by starting at the most complex point and systematically applying simplifications.

3 Brownian Dynamics

The theory of MD considers systems consisting of particles, each of which exerts forces on the surrounding space, including van der Waals and electrostatic forces (Frenkel 2001; Atkins et al. 2018). Imagining a system of N particles, and a particle i with position $\mathbf{x}_i$, we can add up the forces at that point caused by all other particles in the system to get $\mathbf{F}_i(\mathbf{x}_i)$, which is the force that would be felt by particle i.

It is worth going into some detail about what is meant by "particle" in MD. In the most complex forms of MD, "particle" means "atom" (McCammon et al. 1977; MacKerell et al. 1998; Buch et al. 2010). Atoms which are bonded together into molecules experience spring-like forces between each other, ensuring that their average separation is equal to the empirically calculated bond length. Alternatively, at the simplest level, each molecule is considered a single particle, and so forces act on the centre of mass of the molecule rather than its constituent atoms (Alder and Wainwright 1959, 1960). Naturally, there are no spring-like forces in this description since there are

no chemical bonds between particles. In between these two extremes are descriptions of intermediate complexity, where each particle represents a handful of atoms. For instance, one or more water molecules could be treated as a single particle (Wang et al. 2009; Riniker and van Gunsteren 2011), or a protein molecule could be treated as a chain of particles (up to amino acids) connected by spring-like forces (bonds) (Smith and Hall 2001; Ding et al. 2005; Tozzini 2005). In the remainder of this discussion we will, for simplicity, assume the simplest form of MD: "particle" means "molecule". Since all forces act on molecule centres of mass, issues such as molecular shape, molecular orientation and rotation are all absorbed into the force term $\mathbf{F}_i(\mathbf{x}_i)$. For all intents and purposes, it may be simpler henceforth to think of molecules as spheres, though all our arguments apply to general molecular shapes.

It is also worth saying something initially about how chemical reactions occur in MD. Reactions are typically beyond the capabilities of MD simulators: the reasons for this are complex, but fundamentally, it is because reactions are quantum mechanical (Atkins et al. 2018), while MD uses classical mechanics. There are quantum mechanical versions of MD, but they are well beyond the scope of this article and, besides, are extremely computationally intensive (Hu and Yang 2008). We will therefore postpone discussion of chemical reactions until a later point in this section.

We will now consider what we can say about the dynamics of our system of N particles. By Newton's second law, the change in momentum of particle i is equal to $\mathbf{F}_i(\mathbf{x}_i)$. We can therefore write:

$$m_i \frac{\partial^2 \mathbf{x}_i}{\partial t^2} = \mathbf{F}_i(\mathbf{x}_i), \tag{40}$$

where m_i is particle i's mass. This model is perfectly good in theory, but there is an issue of scale. A system of N particles consists of N coupled versions of Eq. (40), each of which is really 3 equations (assuming a three-dimensional system volume), so we end up with $3N$ coupled equations. In the vast majority of systems of interest to biologists, water molecules are by far the most numerous particles: a simple calculation shows that a typical $1\mu\text{m}^3$ *Eschericia coli* cell contains upwards of 10^{10} molecules of water alone. Trying to solve Eq. (40) with N on that kind of scale is simply impossible, even for the best computers. One solution to this is to study only a very small subvolume, much smaller than a cell, possibly containing only one protein (Klepeis et al. 2009). But unfortunately the most interesting biochemistry concerns interactions between several different biological molecules, possibly located in several different parts of the cell, and systems of this kind are beyond the capabilities of molecular dynamics.

The issue, in essence, is that in order to simulate a useful number of interesting molecules a much greater number of water molecules must also be simulated. It was realised that a way around this problem might be to not simulate the water molecules explicitly, but rather to approximate the effect of water molecules on the other molecules by adding terms to Eq. (40) (Pastor et al. 1988). This would dramatically reduce N, and so dramatically increase the size of systems we could plausibly simulate.

The approximation of the water molecules has three components, which have an intuitive justification. We will use the word "collide" to refer to a steric (van der Waals)
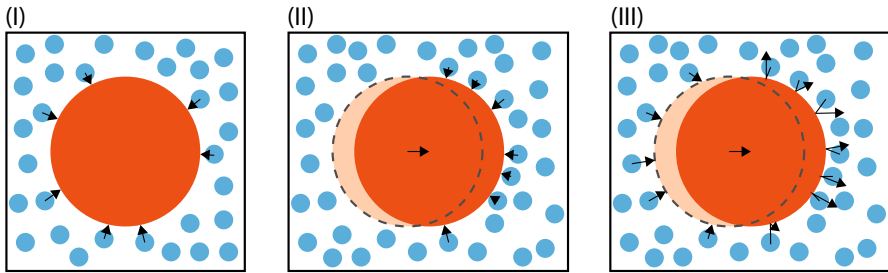
Fig. 7 Three effects of water molecules (blue) on the motion of a large particle X (orange). **I** Water molecules collide on all sides of the large particle apparently at random (black arrows). If X is stationary, there is no directional bias to the random collisions. **II** If X is moving right, water molecules will collide more frequently on the particle's right side and less frequently on its left. There will be a deterministic force opposing X's velocity. **III** The arrows show how the water molecules might move in the time it would have taken them to collide with X had X been stationary (dotted circle). The water molecules on the left have not yet collided with anything, while the molecules on the right have collided and reflected away with higher speeds. They are all further to the right than they would have been otherwise, causing a net flow of water to the right

interaction between particles, this can be thought of as a usual collision between hard objects (or, for simplicity, hard spheres). A graphical explanation of these effects is shown in Fig. 7

(I) When water molecules collide with a larger molecule, X, they will apply a force and thereby induce a change in X's velocity. This is simply due to the conservation of momentum and the conservation of energy. The direction and magnitude of this change are highly variable and are well approximated by a random quantity. Because of the huge number of collisions, the central limit theorem implies that this random force will be approximately Gaussian (Loeve 1977). See Fig. 7 (I).

(II) When X moves with some nonzero velocity, it will tend to collide more frequently with water molecules in its direction of travel. That is, if X is moving to the right, it will tend to experience more collisions on its right side, and fewer on its left side, than the average predicted by (I). The number of collisions will scale with the magnitude of X's velocity (since faster-moving particles will experience more such collisions in a fixed time period). See Fig. 7 (II).

(III) This component is much more subtle and takes some explaining. Let us say X is initially stationary and is hit on the left by a water molecule. Now X starts to move to the right with a fixed speed. After a short time, before it collides with anything else, it will have moved a short distance. Now we think about a potential water molecule which might be about to collide with X. If it is about to collide from above, below, the front or the back, then the fact that X has moved makes essentially no difference to its behaviour. If it is about to collide from the left, it will have to travel slightly further before it collides than it would have if X had not moved. If it is about to collide from the right, it will have to travel slightly less far before it collides. Now suppose the water molecule has collided, and been reflected, and is now travelling away from X. *On average*, the water molecule will be slightly further to the right than it would have been if X had

been stationary. Furthermore, a post-collision water molecule moving right will be moving faster than it otherwise would have been, thanks to the extra energy from the moving X; similarly, a post-collision left-moving water molecule will be moving more slowly. Now imagine another large stationary molecule, Y, to the right of X. It is now slightly more likely to be hit on the left and by a faster-moving water molecule. Another large stationary molecule, Z, to the left of X, is slightly less likely to be hit on the right, and if it is hit it will likely be by a slower-moving water molecule. These two molecules will now both be slightly more likely to move right. Large molecules above, below, in front of and behind X will tend to be unaffected.

Perhaps a more intuitive—though less microscopic—way to think of this is in terms of pressure. If X moves slightly to the right, it will vacate some volume to its left, so there will be a slightly lower water pressure to its left, and water molecules will rush to fill the void. It will also occupy some volume to its right, so there will be a slightly higher water pressure to its right, and water molecules will be pushed out. There will therefore be a net flow of water molecules to the right in the vicinity of X, and so Y and Z will tend to be moved slightly right by the flow.

Mathematically, this manifests itself in a peculiar way: when effect (I) or (II) induces a force on a particle j, effect (III) ensures that all other particles in the system also experience a small force as a result. These forces will be separation dependent, so that two nearby particles will experience strongly correlated forces, while distant particles are weakly correlated. This effect is known as a "hydrodynamic" interaction between particles, because it is really a fluid dynamical effect (Ermak and McCammon 1978).

(I), (II) and (III) all refer to forces induced by steric interactions (collisions) between water molecules and the larger molecule i, but (I) is a random component, (II) is a deterministic component, and (III) is the consequence experienced by molecule i of components (I) and (II) acting on all other molecules in the system. The new version of Eq. (40), implementing all the components, is known as the Langevin equation, and it has the following form:

$$m_i \frac{\partial^2 \mathbf{x}_i}{\partial t^2} = \mathbf{F}_i(\mathbf{x}_i) + \sum_{j=1}^{N} \sqrt{2k_\mathrm{B}T} \left(\Gamma^{\frac{1}{2}} \right)_{ij} \xi_j - \sum_{j=1}^{N} \Gamma_{ij} \frac{\partial \mathbf{x}_j}{\partial t}, \qquad (41)$$

where k_B is the Boltzmann constant, T is the temperature, ξ_i is a standard Gaussian random vector, and Γ is a $3N \times 3N$ matrix known as the "friction tensor". The matrix Γ can be split into 3×3 blocks: the term Γ_{ij} refers to the 3×3 submatrix corresponding to the effect of particle j's motion on particle i. The matrix $\Gamma^{\frac{1}{2}}$ is the square root of the friction tensor.

Effect (I) is described by the term $\sqrt{2k_\mathrm{B}T} \left(\Gamma^{\frac{1}{2}} \right)_{ii} \xi_i$, i.e. the component of the square root friction tensor for the effect of particle i on itself. Similarly, effect (II) is described by $-\Gamma_{ii} \frac{\partial \mathbf{x}_i}{\partial t}$. Effect (III) is incorporated in two distinct ways: first, any random force felt by any particle j, ξ_j, has a hydrodynamic effect on particle i—this is

the reason for the first sum in Eq. (41); secondly, the velocity of any particle j induces a hydrodynamic force on particle i—this is the reason for the second sum in Eq. (41). The hydrodynamic friction tensor Γ mediates these two hydrodynamic contributions, and it is a complicated function of the relative positions of all the particles in the system. The details of the hydrodynamic implementations are well beyond the scope of this article; for a nice summary, see Ref. Ermak and McCammon (1978) which gives examples of the Oseen (Yamakawa 1971) and Rotne–Prager (Rotne and Prager 1969) implementations.

The hydrodynamic effect (III) is often left out of the Langevin equation, because it is subtle and complicated to implement. In that case, instead of Eq. (41) we get the much simpler form:

$$m_i \frac{\partial^2 \mathbf{x}_i}{\partial t^2} = \mathbf{F}_i(\mathbf{x}_i) + \sqrt{2\gamma_i k_B T}\xi_i - \gamma_i \frac{\partial \mathbf{x}_i}{\partial t}, \qquad (42)$$

where ξ_i is now an uncorrelated Gaussian random vector, and the scalar $\gamma_i = \Gamma_{ii}$ is the "friction coefficient". It is easy to see the differential contributions of effects (I) and (II) in Eq. (42). However, there is no physically valid reason to eliminate the hydrodynamic effects at this stage, so we will leave it in for now and discuss when it may be appropriate to remove it later.

Equation (41) is a significant improvement on Eq. (40), and for computational simulations, it does well, but the presence of second derivatives makes it somewhat difficult to study analytically. However, a simplification can be made due to the fact that molecules in water exist in low-Reynolds number conditions, which implies that viscous effects dominate and inertial forces are negligible (Purcell 2014). The result is that $\frac{\partial^2 \mathbf{x}_i}{\partial t^2} = 0$ and so Eq. (41) is simplified.

This is quite complicated, so it worth exploring what is actually happening. Let us temporarily ignore the random collisions with water molecules and think about what happens to a molecule moving through water at a constant velocity. Intuition is not very helpful here, since we are used to thinking about water from a human point of view: if a human is swimming in water at a constant speed, then stops swimming, their motion will slow until they stop completely, i.e. there is a (relatively) lengthy period of decreasing speed (acceleration). This is not what happens to a molecule (e.g. a protein) in water. A closer analogy would be a human swimming through a viscous fluid, such as treacle: when the human stops swimming, they simply stop moving instantly, there is no notable period of decreasing speed. (Of course, there is one, but it happens over such a short timescale that it is negligible.) Water from the point of view of a protein is just like this: it stops moving instantly, there is no protracted period of slowing down. This is the intuitive reason why we can set the acceleration term to zero in Eq. (41). See Ref. Purcell (2014) for an interesting discussion of this effect.

It follows that Eq. (41) becomes:

$$\mathbf{0} = \mathbf{F}_i(\mathbf{x}_i) + \sum_{j=1}^{N} \sqrt{2k_B T}\left(\Gamma^{\frac{1}{2}}\right)_{ij}\xi_j - \sum_{j=1}^{N} \Gamma_{ij}\frac{\partial \mathbf{x}_j}{\partial t}. \qquad (43)$$

Using linear algebra techniques, this set of equations (for $i = 1, \ldots, N$) can be rearranged to give expressions for the velocity of each particle:

$$\frac{\partial \mathbf{x}_i}{\partial t} = \sum_{j=1}^{N} \left(\Gamma^{-1}\right)_{ij} \mathbf{F}_j(\mathbf{x}_j) + \sqrt{\frac{2k_B T}{\gamma_i}} \epsilon_i, \qquad (44)$$

where Γ^{-1} is the matrix inverse of the hydrodynamic friction tensor Γ, and ϵ_i is a standard Gaussian random vector which (unlike ξ_i) is *correlated* with ϵ_j (for all $j = 1, \ldots, N$) according to a complicated hydrodynamic function (Ermak and McCammon 1978). The effect of the linear algebraic rearrangement is to replace the sum of *velocities* in Eq. (43) with a sum of *forces* in Eq. (44). The implication is that any force between two particles (such as a steric repulsion) will induce a (typically miniscule) change in the velocity of every other particle in the system, even if the force does not affect them directly.

Equation (44) is known as Brownian dynamics (Einstein 1956), and in essence, it says that the large molecule's velocity has a deterministic component (due to the forces between each pair of large molecules) and a random component due to random collisions with the water molecules. It is tempting to think that individual collisions with the water molecules induce a random change in the large molecule's position, and that is what Eq. (44) seems to imply, but this is not quite correct. The individual water molecules actually induce random changes in the large molecule's *velocity*, and these changes will tend to happen extremely frequently, so in reality several water molecules contribute to each substantive change in the large molecule's position. This may seem like a semantic distinction, but it is worth bearing in mind that each random position increment corresponds to a large number of actual collisions.

The next simplifying step we can make is to assume that $\mathbf{F}_i(\mathbf{x}_i) = \mathbf{0}$ everywhere and to decorrelate the random vectors ϵ_i, which essentially amounts to ignoring steric (van der Waals), electrostatic and hydrodynamic interactions. This may seem to be a preposterous approximation to make, given that Eq. (44) is already quite simple, but there is a logic behind it. If we were trying to simulate (i.e. solve) Eq. (44) on a computer, we would most likely use an Euler scheme such as the one used in Ref. Ermak and McCammon (1978). To do this, we would pick a small time-step Δt and apply the update rule:

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \sum_{j=1}^{N} \left(\Gamma^{-1}\right)_{ij} \mathbf{F}_j(\mathbf{x}_j)\Delta t + \sqrt{\frac{2k_B T}{\gamma_i}} \epsilon_i \Delta t \qquad (45)$$

to each particle at each time-step. This is not too computationally intensive in itself, but the problem is that for each particle the associated force function $\mathbf{F}_i(\cdot)$, the matrix Γ^{-1} and the size of the correlation in ϵ_i depend on the relative locations of all the other particles in the system. All the different $\mathbf{F}_i$'s, $\left(\Gamma^{-1}\right)_{ij}$'s and correlations must be re-evaluated at every time-step, and this is what seriously slows down the computation. So it is clearly going to be useful to assume $\mathbf{F}_i(\mathbf{x}_i) = \mathbf{0}$, and the ϵ_i's are uncorrelated, but how can we justify this physically? If the concentrations of molecules are

quite dilute, so that pairs of molecules rarely come close enough to interact sterically, electrostatically or hydrodynamically, then $\mathbf{F}_i(\mathbf{x}_i)$ would genuinely be equal to zero and ϵ_i's would be genuinely uncorrelated, for the vast majority of particles, the vast majority of the time.

We can then use the following equation for BD:

$$\mathbf{x}_i = \sqrt{2D_i}\mathbf{W}_t, \tag{46}$$

where $\mathbf{W}_t$ is a three-dimensional Wiener process (confusingly also known as Brownian motion) and $D_i = \frac{k_B T}{\gamma_i}$ is the diffusion coefficient of particle i. The Wiener process, $\mathbf{W}_t = (W_t^{(1)}, W_t^{(2)}, W_t^{(3)})^T$, is a stochastic process defined by making $W_{t+\Delta t}^{(i)}$ a Normal$\left(W_t^{(i)}, \Delta t\right)$ random variable, for $i = 1, 2, 3$. Mathematically, $\mathbf{W}_t$ is the time integral of ξ (Karlin 2014).

It is worth considering the assumption of diluteness that lies behind Eq. (46). Until very recently, and to a large extent currently, diluteness was the *sine qua non* of modelling biochemistry. This was partly due to the modeller's preference for simple models like the REs and the RDEs, which implicitly assume diluteness, but also due to cell biologists' incomplete knowledge about the cellular environment. Then seminal work by Zimmerman and Trach (1991) and Zimmerman and Minton (1993) introduced both modellers and cell biologists to the idea of "macromolecular crowding" (Ellis 2001), an idea which has since become extremely fashionable in fields ranging from computational physics (Torquato and Stillinger 2010; Hofling and Franosch 2013) to bioengineering (Tan et al. 2013; Chapanian et al. 2014).

The basic idea is that the cell contains high concentrations of large molecules. Zimmerman estimated that up to 30% of the internal volume of an *Eschericia coli* cell could actually be occupied by large molecules rather than water (Zimmerman and Trach 1991). Under these conditions, the behaviour of an individual molecule will be seriously affected by the other large molecules in its vicinity, even if they do not interact chemically. There is an open question of how we might accurately modify Eq. (46) to account for this effect. Perhaps the obvious answer is to take a step backwards and bring the $\mathbf{F}(\cdot)$ back into the equation, and this is a common approach (McGuffee and Elcock 2010; Bruna and Chapman 2012; Smith and Grima 2017a), but then we again have the problem of computational efficiency.

There are two other modifications to Eq. (46) which are currently in common usage. The first is to replace the diffusion coefficient D with a modified diffusion coefficient $\tilde{D}$, typically $\tilde{D} = D(1-\alpha\phi)$, where ϕ is the local fraction of volume occupied by large molecules and α is a constant (Weissberg 1963; Blum et al. 1989; Fanelli and McKane 2010; Galanti et al. 2014; Smith et al. 2017). The rationale is that diffusing through a crowded medium might be analogous to diffusing through a viscous medium, and so reducing the diffusion coefficient accounts for an increase in viscosity. There is some computational evidence to back this up, but insufficient for the matter to be considered solved (Saxton 1994; Klann et al. 2009).

Many scientists believe the effect of crowding is more complex than a modification of the diffusion coefficient, and so the second modification is correspondingly less simple. The idea is to replace the Wiener process $\mathbf{W}_t$ with another process with

non-Gaussian increments. The rationale is that the variance of a Wiener process is proportional to t, and some experiments appear to show that the variance of diffusion in a crowded environment is proportional more generally to t^α, an effect known as "anomalous diffusion". Generally, it is observed that $\alpha < 1$ inside cells, which is known as "subdiffusion" (Weiss et al. 2004; Banks and Fradin 2005); however, $\alpha \approx 1$ ("diffusion") (Dauty and Verkman 2004) and $\alpha > 1$ ("superdiffusion") (Reverey et al. 2015) have also been observed. Variants of the Wiener process for pre-specified values of α are commonly used in simulations (Metzler and Klafter 2000; Marquez-Lago et al. 2012), but the idea behind this is not uncontroversial and it raises questions (Dix and Verkman 2008). For example, how do we know which value of α to pick in any given simulation? The debate around anomalous diffusion in the cytoplasm is still very much open, and much more experimental evidence will be needed to reach a satisfactory conclusion.

The question of how to modify Eq. (46) to include hydrodynamic effects is a difficult one and has not been studied in significant detail. It is common to either just use Eq. (44) and live with the computational cost (Ermak and McCammon 1978), or else to simply pretend hydrodynamic effects do not exist (Cichocki and Hinsen 1990; Ridgway et al. 2008; McGuffee and Elcock 2010). The latter option is surprisingly popular and can have significant negative results. Hydrodynamic interactions are believed to be amongst the most important kinds of interactions between biological molecules (Ando and Skolnick 2010), more so than electrostatic effects, and so simply ignoring them cannot be wise. There is some evidence that steric effects and hydrodynamic effects could simply cancel each other out (Felderhof 1978)—so potentially including both or neither in a model might be OK, while including just one might be a bad idea—but much more evidence is needed for this to be a practical solution. Alternatively, some authors have attempted to calculate the effect of hydrodynamic interactions on diffusion coefficients (Felderhof 1978; Ohtsuki and Okano 1982) and reaction rates (Deutch and Felderhof 1973)—then it might be possible to modify them accordingly and claim that hydrodynamics have been taken into account. This is probably the most plausible approach, but again considerably more work is needed to make this method practical.

To bring the discussion back to Brownian dynamics, there is still something missing from our model Eq. (46): How do reactions happen? Reactions with 1 reactant (i.e. unbinding reactions) are easy: each molecule which can unbind has an internal clock, and the time until unbinding is generated as an exponentially distributed random time in the future—when the clock exceeds this time, the reaction occurs. There is a question of where the daughter molecules (i.e. the products of the unbinding) should be placed, and several different methods are in use. Perhaps the simplest method is to place them both at the same location, where the parent particle was before it unbound: this is unphysical but straightforward and requires no new parameters (Lipkov et al. 2011). An alternative is to propose an "unbinding distance", σ, and place them randomly opposite each other on a sphere of diameter σ centred around the location of the parent particle (Andrews and Bray 2004).

On the other hand, modelling *bimolecular* reactions is a complicated question with a very long history. The original method, still popular today, was devised by Smoluchowski (1917), who incidentally was one of the originators of the physical theory of Brownian motion, along with Albert Einstein (Marian 1924; Einstein 1956).
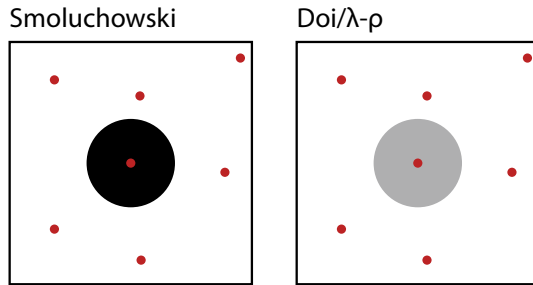
Fig. 8 Two methods of implementing bimolecular reactions in BD. The Smoluchowski model allows particles to react immediately when the distance between the particle centres equals the sum of the particles' reaction radii. This radius of certain interaction is shown as a solid black circle. The Doi/$\lambda - \rho$ model allows particles to react with probability λ per unit time when the distance between the particle centres is less than a distance ρ. This radius of probabilistic interaction is shown as a solid grey circle. Note that the particle centres are denoted in red

Smoluchowski's idea was a simple one: each molecule is assigned a "reaction radius", and two reacting particles will react immediately if they are brought close enough together that their reaction radii overlap (see Fig. 8). This model has advantages and disadvantages.

The advantages are that it is simple to understand and computationally and mathematically very straightforward. The most popular Brownian dynamics software, *Smoldyn* (Andrews and Bray 2004), uses the Smoluchowski model because it is so much faster than the alternatives. Mathematically, it is possible to write down the expected time until a reaction between two particles, μ, as a simple function of their reaction radii (r_1, r_2), their diffusion coefficients (D_1, D_2), and the reaction volume V:

$$\mu = \frac{V}{4\pi (r_1 + r_2)(D_1 + D_2)}. \tag{47}$$

Equation (47) is obtained from the diffusion equation, and a derivation can be found in Ref. Gillespie (2009a). In principle, μ^{-1} could be put into the CME as the rate of a reaction, although this would be an approximation because the actual waiting time for Brownian collisions is not exponential (Redner 2001).

The disadvantages of Smoluchowski's approach are slightly more complicated. Principally, the Smoluchowski model is not a particularly accurate model of how bimolecular reactions actually happen. Reacting molecules must approach each other in the correct orientation and with sufficient combined kinetic energy to exceed the activation energy of the reaction; otherwise, they will simply collide without reacting (Atkins et al. 2018). Smoluchowski does not allow for these kinds of collisions, which can occur much more frequently than successful reactions.

The second issue is a modelling issue rather than a physical one. We have already discussed the concept of "infinite diffusion", which is unphysical but extremely useful for modelling purposes. Our ideal BD model is one where an infinite diffusion coefficient does not give nonsensical results. In the Smoluchowski model, the expected time until a reaction scales as the inverse of the diffusion coefficient, and so when dif-

fusion is infinitely fast bimolecular reactions happen infinitely quickly. It may appear that Smoluchowski has it right here because we know from basic chemistry that if we increase the temperature of a system (and therefore the diffusion coefficients) the reactions will tend to occur more quickly (Atkins et al. 2018). But the reality is not quite so simple.

A significant source of confusion arises from the fact that a generalised bimolecular reaction rate k (such as the rate in the REs) really incorporates two distinct processes: a diffusive process which brings two molecules together, and a reactive process which causes them to react with some probability. Collins and Kimball go into considerable detail on this subject in Ref. Collins and Kimball (1949), as well as Gillespie later in Ref. Gillespie (2009a). They consider a physically plausible system in which hard sphere particles diffuse by Brownian motion, but over very short timescales (shorter than the time between collisions with solvent molecules) they have instantaneous random velocities given by the Maxwell–Boltzmann distribution. In this description, a reaction occurs when the instantaneous velocities of a very close pair of particles combine in such a way that they collide with an energy greater than the intrinsic activation energy of the reaction. Under these conditions, Collins and Kimball observe that the reaction rate k satisfies:

$$\frac{1}{k} = \frac{1}{k_d} + \frac{1}{k_b},$$ (48)

where k_d is the *diffusive rate* and k_b is the *ballistic rate*. In other words, the expected time until a reaction between a given pair of particles is the sum of the expected time until they are brought together by diffusion and the expected time until they subsequently collide with sufficient energy.

According to Collins' and Kimball's model (1949), the role of diffusion is to bring reactive molecules together, whereupon they can react according to some diffusion-independent process. In other words, the probability of a reaction between a sufficiently close pair of particles (in a sufficiently short time that they cannot diffuse away again) is not a function of the diffusion coefficient. When the diffusion coefficient is infinitely fast the time until a reaction, $\frac{1}{k}$, approaches $\frac{1}{k_b}$ rather than 0, and so infinite diffusion seems not to pose a problem. Unfortunately, again, it is not so simple.

Since Collins' and Kimball's k_d is a linear function of the diffusion coefficient D, while k_b has no D-dependence, our assumption may seem to be straightforwardly valid. However, both D and k_b are functions of temperature, via the Stokes–Einstein relation and the Arrhenius equation, respectively (Atkins et al. 2018). It follows that k_b has an implicit D-dependence—e.g. if D is large, there is a possibility that this may be due to the temperature being large, so k_b may also be large. In biology (unlike chemistry) this is not a serious problem, since temperature is essentially constant for living beings. Our assumption amounts to ignoring this subtle implicit dependence, and thereby restricting our models to biological systems. We assume that it is possible to freely vary D without affecting the reactive component of the bimolecular reaction rate—we can think of this as varying D by tuning the viscosity of the solvent, which does not directly affect the reaction rate.

In this spirit, a popular alternative to the Smoluchowski model was proposed by Doi (1976), building on the work of Teramoto and Shigesada (1967). He suggests that

we should associate with each pair of reacting particles a reaction rate λ and a reaction distance ρ, then we assume again that waiting times between reactions are exponential, so that two particles will react in a short time-step Δt with a probability $\lambda \Delta t$ if they are separated by a distance less than ρ. In other words, if two particles are sufficiently close they react with a rate λ; otherwise, they do not react (see Fig. 8). In practice, at each time-step Δt particle positions are updated according to Eq. (45) (typically minus the force term), whereupon the distances between each pair of potentially reacting particles are calculated. For each pair, if this distance is less than the ρ associated with the reaction, then the reaction occurs with probability $\lambda \Delta t$, the particles are removed from the system, and the product particles (if there are any) are put in their place.

Doi's model, which has come to be known as the $\lambda - \rho$ model (Erban and Chapman 2009), is actually a generalisation of the Smoluchowski model, which would have $\lambda = \infty$ and ρ equal to the sum of the reactants' reaction radii (Agbanusi and Isaacson 2014). Overall, the $\lambda - \rho$ model seems to be an optimal trade-off between model simplicity and physical accuracy. The process we want to model explicitly (diffusion) is modelled explicitly, while more complex processes (reactant orientation, kinetic energy, etc.) are lumped into the rate λ. The main disadvantage is that we are assuming exponential waiting times between reactions, but this cannot really be helped without going into significantly more modelling detail and thereby incurring a significantly higher computational cost.

We have seen how the $\lambda - \rho$ model was obtained from a purely Newtonian molecular dynamics model by making a variety of assumptions and simplifications, so that unlike the RDME, $\lambda - \rho$ is a *bottom-up* model. This provides a certain guarantee about physical accuracy, and so $\lambda - \rho$ is frequently used as a "ground truth" model. However, there is still the question of how $\lambda - \rho$ fits in with the simpler models on the complexity scale: the CME, the RDEs and the REs. We therefore now analyse the limiting behaviour of $\lambda - \rho$ in the limits of fast diffusion and high concentrations

3.1 The Limit of Fast Diffusion

To analyse the limit of fast diffusion, we assume that diffusion coefficients can be varied independently of the λ's and ρ's. We particularly focus our attention on the limiting case where all diffusion coefficients tend to infinity (the "reaction-limited" case) and particles cannot simply diffuse out of the reaction volume. (We would say that the boundaries of the volume are "reflective".) In this case, the position of a particle at a time $t + \Delta t$ is a uniform random variable independent of its position at time t. To see this, note that the updated position of a particle conditioned on its previous position will be a Gaussian random variable with large variance; the boundary conditions cause this distribution to approximate a uniform distribution, and the approximation improves as the variance grows.

Now, consider a reaction of the form $A + B \rightarrow \emptyset$, with rate λ and reaction distance ρ. The symbol $\emptyset$ denotes that we are not interested in the products of the reaction. The probability that a reaction occurs between a given pair of A and B molecules in a time period $[t, t + \Delta t)$ is given by:

$$P(\text{reaction in } [t, t + \Delta t)) = K(\rho, [t, t + \Delta t))\lambda\Delta t, \qquad (49)$$

where $K(\rho, [t, t + \Delta t))$ represents the exact proportion of the time interval $[t, t + \Delta t)$ for which the given pairs are within a distance ρ of each other. Typically, K not only is unknown, but also is itself a random variable, so that reactions in the $\lambda - \rho$ model are doubly stochastic processes. In the limiting case where $D \to \infty$, K becomes a known random variable. At any given time point in $[t, t + \Delta t)$, each molecule is uniformly distributed, so that the probability of any pair being within ρ of each other is some constant c. The constant c is approximately equal to $\frac{4\pi\rho^3}{3V}$, the proportion of the total volume occupied by a sphere of radius ρ; c is not exactly equal to this quantity because of boundary effects: if one particle is close to a boundary, there will be less space around it for a second particle to react in. However, these kinds of effects are typically numerically very small, and the approximation is usually excellent.

At any given time point in $[t, t + \Delta t)$, then, the event "A and B are within ρ of each other" is governed by a Bernoulli(c) random variable, say C. If we consider N equally spaced time points in $[t, t + \Delta t)$, then we can approximate:

$$K \approx \frac{1}{\Delta t}\sum_{i=1}^{N} C_i \frac{\Delta t}{N} = \frac{1}{N}\sum_{i=1}^{N} C_i, \qquad (50)$$

where C_i are independent Bernoulli(c) random variables. The approximation converges to equality in the limit $N \to \infty$, so by the law of large numbers, we have that $K(\rho, [t, t + \Delta t)) = c$, so that

$$P(\text{reaction in } [t, t + \Delta t)) = c\lambda\Delta t. \qquad (51)$$

This highlights the remarkable mathematical effects of allowing the limit $D \to \infty$: an unknown random variable K is converted into a known constant c. From Eq. (51) we can derive an evolution equation for $P(n_A, n_B, t)$, the probability that there are n_A molecules of A and n_B molecules of B at time t. We choose Δt sufficiently small that at most one reaction can happen in the time period $[t, t + \Delta t)$, and we observe that the probability of a reaction happening given that we have n_A, n_B molecules of A, B, respectively, is $c\lambda\Delta t n_A n_B$ [since Eq. (51) gives the probability that any given *pair* will react, and there are $n_A n_B$ pairs]. The evolution equation is:

$$P(n_A, n_B, t + \Delta t) = P(n_A + 1, n_B + 1, t)(n_A + 1)(n_B + 1)c\lambda\Delta t$$
$$+ P(n_A, n_B, t)(1 - n_A n_B c\lambda\Delta t), \qquad (52)$$

which in the limit $\Delta t \to 0$ becomes:

$$\frac{\mathrm{d}}{\mathrm{d}t} P(n_A, n_B, t)$$
$$= \lambda c \left[(n_A + 1)(n_B + 1)P(n_A + 1, n_B + 1, t) - n_A n_B P(n_A, n_B, t)\right]. \qquad (53)$$

Fig. 9 The region (orange) within a reaction distance ρ of a central subvolume (green). The volume is subdivided into microscopic subvolumes, with side length substantially smaller than ρ
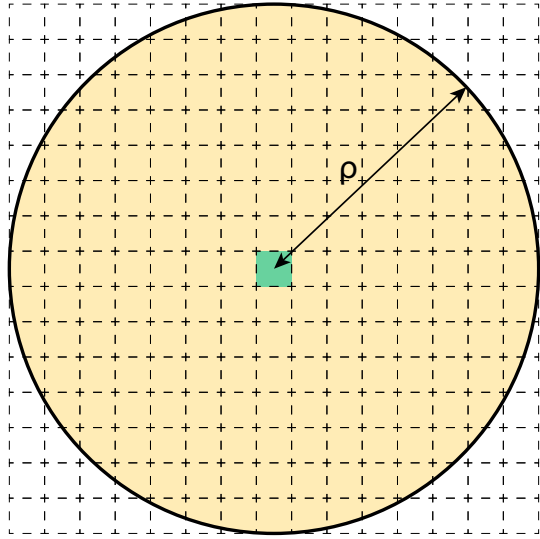
This equation is precisely the CME for the system $A + B \xrightarrow{k} \emptyset$, where $k = \lambda c V$. The CMEs for other types of mass action reactions can be derived in an analogous way, and the same will hold true. It follows that the CME *exactly* describes the evolution of molecule numbers in $\lambda - \rho$ in the limit of fast diffusion coefficients.

The CME has also been derived by Gillespie as the exact description for a system of uniformly distributed molecules which have instantaneous velocities given by the Maxwell–Boltzmann distribution and which react if they collide with sufficient combined velocity (Gillespie 1992). There is question with this derivation as to whether it is possible for a particle to be uniformly distributed in space at all times and also to have a finite velocity and specific location. The issue is that if, say, a pair of particles collide and fail to react, then they are necessarily close together and so much more likely to react than any other pair in the next short time interval. This violates the assumption of uniform distributions and also the principle of mass action. It is our belief that the CME can be considered correct if and only if there is a genuine separation of timescales between particles' motion (i.e. diffusion) and their reaction rates—whether we think of this as "infinite diffusion" or as "reaction-limited" dynamics is essentially a semantic question.

3.2 The Limit of High Concentrations

We next study the behaviour of BD when the concentration of molecules is very high. We will consider a reaction of the form $A + A \rightarrow \emptyset$ with rate λ and reaction distance ρ, though the arguments can be developed analogously for other reaction types. We consider a small subvolume ΔV with a diameter substantially *smaller* than the reaction distance ρ. This set-up is shown in Fig. 9. A particular subvolume is shown in green, a circle with radius ρ is drawn around it (orange), such that any molecule within this

circle can react with any molecule in the green subvolume with rate λ. The dotted squares denote subvolumes of size ΔV.

Since the concentration of molecules is very high everywhere, as long as the diffusion coefficient is nonzero, the expected concentration of molecules is locally constant inside the reaction distance and the variance is negligible. We define ϕ as the local expected concentration, so that the average number of molecules per subvolume is $\phi \Delta V$.

We now consider the expected change in molecule numbers in the central subvolume in a short time Δt. Temporarily ignoring diffusion, the molecule number can change for two reasons: (1) two molecules in the central subvolume (green) react with each other and (2) a molecule in the central subvolume reacts with a molecule outside the central subvolume (orange). The expected change in molecule number due to (1) is $-2\lambda \Delta t$ times the number of pairs in the central subvolume, i.e. $-2\lambda \Delta t \frac{\phi \Delta V (\phi \Delta V - 1)}{2}$. Note that the initial 2 comes from the fact that a single reaction reduces the molecule number by 2. The change in molecule number due to (2) is $-\lambda \Delta t \phi \Delta V$ times the number of molecules outside the central subvolume but inside the reaction radius (orange), i.e. $-\lambda \Delta t \phi \Delta V \phi (V_3(\rho) - \Delta V)$, where $V_3(r) = \frac{4\pi r^3}{3}$ is the volume of a sphere with radius r. The overall expected change is then given by adding the contributions from (1) and (2):

$$- \lambda \Delta t \phi \Delta V \left[\phi V_3(\rho) - 1 \right] \approx - \lambda V_3(\rho) \phi^2 \Delta V \Delta t, \tag{54}$$

where the approximation arises from assuming the concentrations are sufficiently high that $\phi \gg \frac{1}{V_3(\rho)}$.

In the reaction–diffusion equations, the rate of change of concentrations due to the reaction $A + A \rightarrow \emptyset$ at a point x is given by $-2k\phi(x)^2$, where k is the reaction rate and $\phi(x)$ is the local concentration near x. It follows that the expected change in molecule numbers in a small volume ΔV around x in a short time Δt is given by:

$$- 2k\phi(x)^2 \Delta V \Delta t. \tag{55}$$

It follows that the RDEs *exactly* describe the reactive behaviour of BD with high concentrations, if we choose $k = \frac{\lambda V_3(\rho)}{2}$. The corresponding fact that the diffusion equation correctly describes Brownian diffusion is extremely well established (Gillespie and Seitaridou 2012), and we will not go into the subject here.

We note that our justification of the RDEs here implies that a molecule is unlikely to react with another molecule in the same subvolume and is much more likely to react with another molecule in a different, nearby subvolume. This makes intuitive sense when ΔV is very small, but nonetheless runs counter to the usual intuition behind the RDEs. This idea is also the basis for the derivation of the CRDME (Isaacson 2013).

Overall, we have seen that the $\lambda - \rho$ implementation of BD is obtained from MD in a relatively rigorous manner (Erban 2014), by systematically simplifying out various effects like water and inertia. Subsequently, we found that, like the RDME, $\lambda - \rho$ can be thought of as a "parent model" of the CME, the RDEs and the REs. The relationship between all of these models is shown in Fig. 10. Just as with the RDME, if we use both $\lambda - \rho$ and the RDEs (for example) to model the same system, and the system has parameters such that we expect the RDEs to be correct, then the two models will give
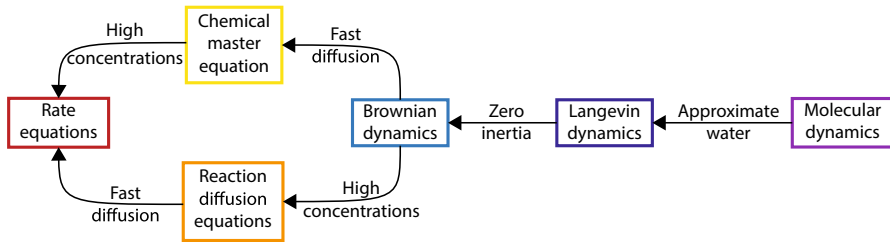
Fig. 10 How the $\lambda - \rho$ BD model fits into the scale of model complexity shown in Fig. 1. It is the midpoint of complexity between MD and the REs

identical predictions. This clearly demonstrates BD's place in the complexity scale (Fig. 1), and as a model with a sound microphysical basis and clear connections to other mainstream models, it is easy to see why BD is such a popular technique. It appears, then, that BD and the RDME have a lot in common: in the next section, we briefly assess the extent to which the two can be said to agree.

4 Does the Reaction–Diffusion Master Equation Agree with Brownian Dynamics?

The $\lambda - \rho$ model has a lot in common with the RDME. Both models consider diffusing point-particles which can react with a fixed probability only if they are sufficiently close together. The rate λ is analogous to the scaled bimolecular rate $\frac{kM}{V}$, and the subvolume size $\frac{V}{M}$ is analogous to the sphere of radius ρ around the Brownian particle. The biggest difference is that the RDME discretises space, but keeps time continuous (reactions can occur at any time, but only in fixed subvolumes), whereas almost all $\lambda - \rho$ simulators discretise time and keep space continuous. (Reactions can occur anywhere, but only at time-steps of pre-specified length.) The exception is the already-mentioned CRDME (Isaacson 2013), which can be thought of as a discrete-space continuous-time $\lambda - \rho$ simulator.

It might then be surprising to learn that there is no established connection between the RDME and $\lambda - \rho$. Given a particular implementation of the $\lambda - \rho$ model for a particular system, it is typically possible to find values of M and reaction rates such that the RDME agrees by some metric, but if some aspect of the system (e.g. a diffusion coefficient) is changed even slightly, the optimal RDME parameters will typically have to be found again. This fact was demonstrated neatly in Ref. Smith et al. (2016). An analytical approximation to the mean molecule numbers in the RDME was obtained, which is a function of both M and the diffusion coefficient D. For any fixed choice of M, it was shown that this function agrees with $\lambda - \rho$ only for a specific choice of D. Correspondingly, for any fixed value of D, only a specific choice of M gives reasonable agreement with $\lambda - \rho$.

To help understand this issue better, we performed simulations of the system studied in Ref. Smith et al. (2016) using both the RDME and $\lambda - \rho$. The system is straightforward and consists of a single species A and two reactions:
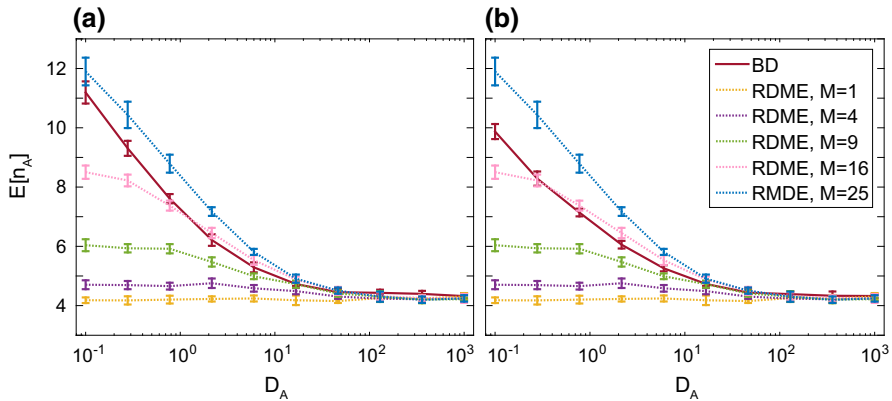
Fig. 11 Comparison of RDME with $\lambda - \rho$ BD for the system (56), showing the dependence of the average number of A molecules, $\mathbb{E}[n_A]$, as a function of diffusion coefficient, D_A. **a** BD corresponds to $\lambda - \rho$ with spherical particles. **b** BD corresponds to $\lambda - \rho$ with cubic particles. Error bars are the standard deviation of 10 independent SSA simulations. Parameter values are $k_1 = 1000$, $k_2 = 30$, $\rho = 0.1$, $\lambda = 1900$, $V = 1$. Simulations are performed in two spatial dimensions

$$\emptyset \xrightarrow{k_1} A, \ A + A \xrightarrow{k_2} \emptyset, \tag{56}$$

where $\emptyset$ denotes that we are not interested in the species involved. The first reaction is assumed to occur at uniformly distributed points in space: it could represent transcription of a protein A where we do not explicitly model mRNA or ribosomes. We simulated system (56) using the RDME with a variety of values of M ranging from 1 to 25, and for a variety of diffusion coefficients D_A ranging from 10^{-1} to 10^3. In Fig. 11a we plot the mean number of A molecules n_A as a function of D_A for each choice of M: the results are quite remarkable.

For each value of M (except $M = 1$) we find that the mean molecule number decreases with diffusion. When diffusion is fast, all RDME simulations agree—this should not surprise us, since we have already proved that they will all converge to the same model (the CME) in this limit. But when diffusion is slow, the RDME for each value of M appears to converge to a *different* limit: when $M = 1$, the limit appears to be around 4, but when $M = 16$ it appears to be around 9. The implication is that the RDME appears to be very sensitive to the choice of M, at least when diffusion is not fast (and when diffusion is fast you may as well use the CME).

Smith et al. (2016) cannot be said to be the first time an effect of this kind was noticed. It is well known (and we have already observed) that bimolecular reactions tend not to happen in the RDME when M is "too large", which would naturally lead to changes in mean concentration as M grows (Isaacson 2013). But this is not quite the same effect as the one shown in Fig. 11a. Bimolecular reactions are happening for every choice of M, they are just happening at slightly different effective rates, with the result that every M behaves differently (not just the "too large" ones). The standard solution—to choose M not "too large"—does not appear to be very helpful here.

This slightly more subtle effect was observed by Erban and Chapman [who were also investigating system (56)], and they devised an ingenious solution (Erban and

Chapman [2009]). They suggested that the value of the bimolecular reaction rate, k_2, should be modified in a manner that depends on M, and derived a formula for this modification. If $M = 1$, k_2 should be left as it is; if $M = 4$, k_2 should be reduced slightly; and if $M = 25$, k_2 should be reduced significantly. The result is a rescaled RDME whose results do not depend on M. In our Fig. 11a, this would lead to every RDME curve looking identical to the orange ($M = 1$) curve, i.e. there is no dependence on M or D_A. This appears to solve the problem: we can now freely vary M (up to a point, remember if M is "too large" bimolecular reactions will simply not happen) without worrying about the choice of M impacting on our results.

There is a possibility, however, that negates the neat solution of Erban and Chapman: What if the mean molecule numbers depend on D_A for a good (i.e. physical) reason? If the choice of D_A actually affects system (56) in a significant way, and we modify the RDME so that it always agrees with a model which does not depend on D_A (the CME), then surely we are incorrectly modelling the system. There is a simple way to check whether the D_A-dependence is a real effect or an artefact of the RDME, and that is to simulate it with $\lambda - \rho$.

The solid burgundy line in Fig. 11a corresponds to an average of $\lambda - \rho$ simulations, with λ and ρ chosen to agree with the CME when D_A is large. (This is always possible, as described in Sect. 3.1.) We find the same effect: the mean molecule number decreases with D_A, implying that this change is not an RDME artefact, but a real effect which is important to model correctly. This brings us back to our original problem: Which value of M should we pick? The surprising implication of Fig. 11a is that there is no definitive value of M. If D_A is large, any value of M will agree with $\lambda - \rho$. In the range $10^1 < D_A < 10^2$, it appears that $M = 9$ gives the best agreement, but for $10^0 < D_A < 10^1$, the choice of $M = 16$ is optimal. As D_A gets lower, the optimal value of M gets larger. This leads us to one of the main results of Smith et al. ([2016]): for any choice of M (say, $M = 9$), only a subset of values of D_A are accurately modelled by the RDME ($D_A > 10^1$), and for any choice of D_A (say, $D_A = 2$), only a particular value of M will be optimal ($M = 16$).

Optimally choosing M in the RDME is a difficult task, and several authors have proposed methods to solve the problem based on physical arguments (Bernstein [2005]; Isaacson and Peskin [2006]; Bayati et al. [2011]; Kang et al. [2012]). The most common approach is to choose M such that the diffusive timescale is substantially shorter than the reactive timescale, i.e. in a given short time-step Δt, a molecule is much more likely to diffuse into a new subvolume than it is to react (Bernstein [2005]; Isaacson and Peskin [2006]; Bayati et al. [2011]). This will guarantee that each subvolume is well mixed, and so a CME-like description can be said to be appropriate on the subvolume scale. In the case shown in Fig. 11a, the rate of diffusion between two neighbouring subvolumes is $\frac{D_A}{h^2}$, for lattice spacing h; since we are considering a two-dimensional volume, we have that $h^2 = \frac{V}{M}$ and that a molecule can diffuse in four different directions, so that the probability that a given molecule diffuses in Δt is $\frac{4 D_A M \Delta t}{V}$. Similarly, the rate of a bimolecular reaction in any given subvolume i is $\frac{k_2 M}{V} n_A^{(i)} (n_A^{(i)} - 1)$, where $n_A^{(i)}$ represents the number of molecules of A in subvolume i. The probability that a given molecule is involved in this bimolecular reaction in time-step Δt is then approximately given by $\frac{k_2 M}{V} \mathbb{E}[n_A^{(i)}] \Delta t$, and noting that by spatial uniformity we have $\mathbb{E}[n_A^{(i)}] = \frac{\mathbb{E}[n_A]}{M}$,

we get a final probability of $\frac{k_2}{V}\mathbb{E}[n_A]\Delta t$. Overall, then, we require $M \gg \frac{k_2\mathbb{E}[n_A]}{4D_A}$. In the case where $D_A = 1$, and taking $\mathbf{E}[n_A] \approx 4$ (since this is all we can say from the REs), we find the requirement $M \gg 30$, clearly at odds with Fig. 11a in which $M = 16$ is optimal. However, more importantly, since Fig. 11a shows there is a single value of M which agrees with BD, one-sided bounds could *never* give sufficiently useful information to ensure agreement between BD and RDME. We conclude that only an estimate of the actual value of M (or very tight two-sided bounds on M) could potentially work here; however, it currently seems doubtful that such an estimate could be derived in general.

Again, we see the uneasy relationship between the RDME and $\lambda - \rho$, which both model the same kinds of systems at roughly the same level of complexity and yet never seem to agree definitively about anything. Why is this? A superficially convincing argument is to do with shape: because $\lambda - \rho$ uses the usual L_2 distance metric, a molecule can react with any other molecule within a spherical region around it, but the RDME typically uses cubic subvolumes, so a molecule can react only with other molecules within a cubic region. Could it be that the difference between spheres and cubes is the driver of disparity between the RDME and $\lambda - \rho$? There is a simple way to check. The $\lambda - \rho$ model can equally be defined to use other distance metrics, such as the one which leads to cubic reaction regions around molecules (Opplestrup et al. 2006). A metric of this kind was used to generate the data plotted in Fig. 11b, which compares the "cubic" $\lambda - \rho$ with the same RDME simulations as in Fig. 11a. Exactly the same effect is observed as in Fig. 11a: the optimal M for agreement with $\lambda - \rho$ changes as D_A changes.

If it is not shape related, why then is there this discrepancy between RDME and $\lambda - \rho$? This is a relatively open question, but the most convincing argument is that modelling bimolecular reactions accurately requires fine-grained spatial resolution, something which $\lambda - \rho$ clearly has, and the RDME clearly does not. As evidence for this, consider that the CRDME is a modification of the RDME allowing it to have fine-grained spatial resolution and is proven to be able to approximate $\lambda - \rho$ to any given degree of accuracy (Isaacson 2013).

5 Discussion

In this review, we have concerned ourselves with two stochastic reaction–diffusion models: the RDME and the $\lambda - \rho$ implementation of BD. These two models are very similar in complexity, and they both model the same kinds of processes at similar levels of detail. Indeed, we have seen that they both converge to the CME in the limit of fast diffusion, they both converge to the RDEs in the limit of high concentrations, and they both converge to the REs when both of these limits are applied simultaneously. And yet despite this, there appears to be no direct relationship between the RDME and $\lambda - \rho$, and they typically do not agree numerically.

Since $\lambda - \rho$ is derived in a semi-rigorous manner from an even more fundamental model, it may appear that there is no reason to use the RDME at all and that when $\lambda - \rho$ and RDME disagree, it is because the RDME is simply wrong. This is not quite correct. As we have seen, $\lambda - \rho$ and the RDME were developed to model the same kinds of

systems, but they have completely different historical origins. The RDME is obtained by starting from a simple model (the REs) and working backwards, adding in stochasticity and diffusion—it is therefore a top-down model. $\lambda - \rho$ is obtained by starting from a complex model (MD) and working forwards, removing water molecules, and steric, electrostatic and hydrodynamic interactions—it is a bottom-up model. Remnants of these origins can still be seen in the RDME (the reaction rate k from the REs) and $\lambda - \rho$ (the reaction distance ρ, which is akin to a van der Waals radius from MD).

In an ideal world, where computers were infinitely fast, it may be tempting to use MD for all chemical kinetics problems, but this may not be wise. In many cases, particularly in biology, we do not know all the chemical species involved in a particular system, nor do we know exactly how many molecules of each species there might be, nor the details of the reactions between them. There are a lot of parameters in MD, and consequently, a lot of possible ways to get the model wrong. On the other hand, RE models are very popular, because for each reaction there is only one parameter (the reaction rate k), which can be relatively easily estimated from experimental data. It is hard to get the model *wrong*, and more likely the issue will be that the model does not incorporate some significant effect (e.g. diffusion). It is easy to turn a RE model into a RDME model: the only choice that needs to be made is the number of subvolumes, M. It is hard to turn a RE model into a $\lambda - \rho$ model, because for each bimolecular reaction there are two parameters to choose (λ and ρ), and it is not clear how any particular choice of these parameters will relate to the k's from the REs unless diffusion is very fast.

In many practical cases, the RDME model requires the fewest assumptions, and so will typically be the least wrong model. We have seen that the RDME can give very different numerical predictions depending on the choice of M, but typically all choices of M will agree *qualitatively*, and if numerical precision is not paramount the RDME may be the best choice. In short, a good rule of thumb is this: if you know a lot about the details of your system (e.g. molecular sizes), use $\lambda - \rho$; otherwise, use the RDME.

A further typical advantage of the RDME over BD, which we have not touched upon in this review, is computational speed (Erban et al. 2007). By virtue of its spatial discretisation, the RDME typically models fewer distinct quantities than BD and will tend to be substantially faster when molecule numbers are large and subvolumes are not too small. On the other hand, BD will perform better if RDME subvolumes contain very few molecules on average. In some cases, different models may be appropriate in different regions of the cell, and so a hybrid of BD and RDME may be the optimal model (Flegg et al. 2014). Alternatively, significant speed boosts can be achieved in BD using the Green's function reaction dynamics (GFRD) approach which models Brownian diffusion in a statistically exact manner analogous to the SSA for the CME (van Zon and ten Wolde 2005; Takahashi et al. 2010). The rationale behind GFRD is that if a molecule is far from other molecules there is no point in simulating its entire trajectory of *positions* with a fixed small time-step, and it is more efficient to sample new *times* from the first passage time distribution of the time taken for it to reach a pre-specified displacement from its origin. The speed boost of GFRD relies on the assumption that this sampled time-step will be longer on average than the equivalent time-step in standard BD simulations, which is typically true for dilute systems but not when concentrations are high, making GFRD a poor choice for realistic whole-cell
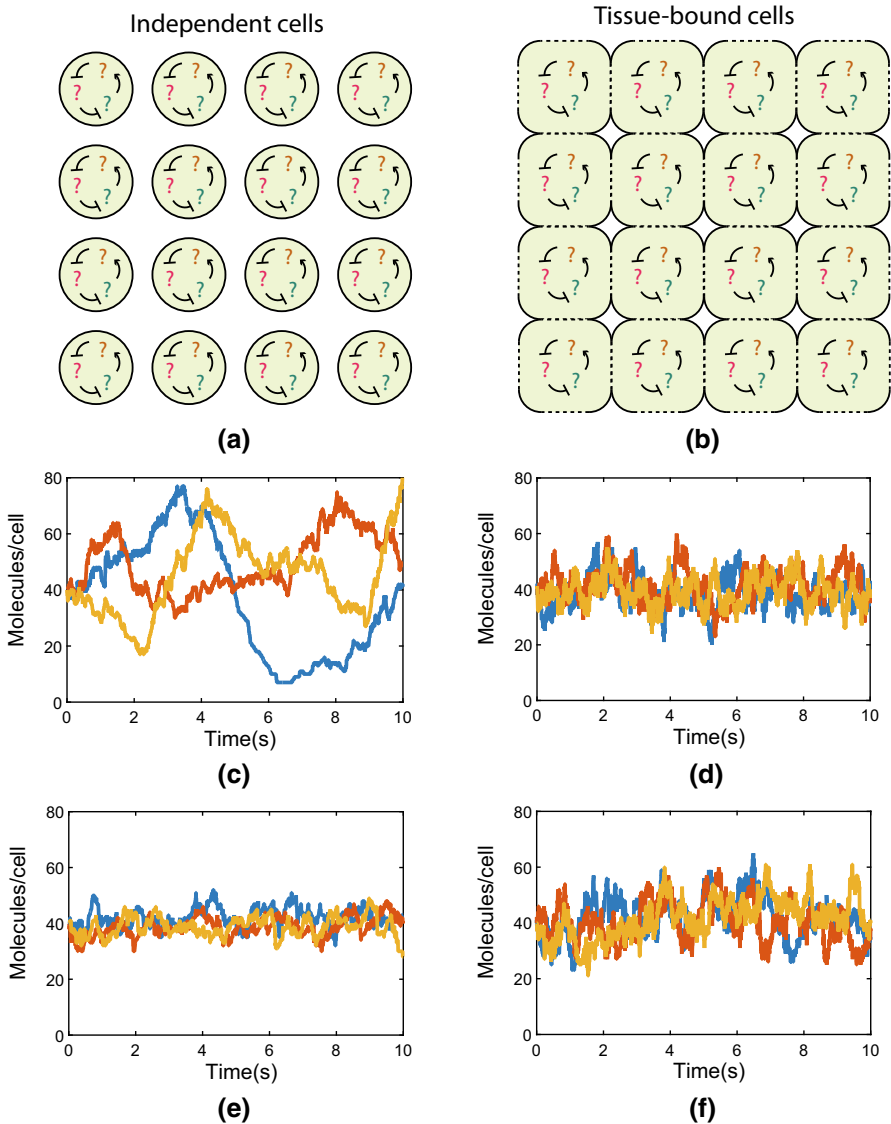
Fig. 12 Effect of molecular transport on tissue heterogeneity. **a** A population of unconnected cells, with no intercellular transport. **b** A tissue of cells, with transport between neighbouring cells. **c** Simulations of a noisy system in unconnected cells: the cell concentrations are very heterogeneous. **d** Simulations of the same system in a connected tissue: the heterogeneity is significantly reduced. **e** Simulations of a less noisy system in unconnected cells: the cell concentrations are quite homogeneous. **f** Simulations of the same system in a connected tissue: the heterogeneity is increased

modelling. Furthermore, GFRD algorithms have only been derived for Smoluchowski reaction kinetics, and not the $\lambda - \rho$ kinetics for which we have argued in this review.

An alternative use of the RDME has recently been proposed, which eliminates the concerns about the choice of M (Twycross et al. 2010; Smith et al. 2016; Smith and Grima 2018). The idea is to consider each subvolume of the RDME as a cell in a

tissue, and "hopping" events correspond to transport (active or passive) of molecules between adjoining cells. In this description, M simply refers to the number of cells in the tissue, and so is no longer an undetermined parameter. An interesting and counter-intuitive result was derived in Ref. Smith and Grima (2018) by thinking along these lines, and generalising a related result originally derived in Ref. Erdmann et al. (2009) for single-celled organisms. It was discovered that transport between cells (the analogue of diffusion in the classic RDME) has a non-trivial relationship with the heterogeneity of the tissue, i.e. the tendency for different cells in the tissue to have different concentrations. If a system has large fluctuations in concentration in a single-cell (i.e. if it is very noisy), molecular transport will tend to make the tissue more homogeneous, but if a system has small fluctuations (i.e. if it is not very noisy) then transport will tend to make the tissue more heterogeneous. This result is completely generic, though a similar result was shown by Erdmann et al. for a specific gene system. The basic principles of this are shown in Fig. 12. The idea of rationalising the RDME in this way is still in its infancy, but it is certainly promising: it could have significant usage in the modelling of cancer, which can develop stochastically in individual cells in a tissue (Gupta et al. 2011). A similar idea has also been used to model single cells by treating each physical compartment of the cell (nucleus, cytosol, etc.) as a separate subvolume of the RDME (Isaacson and Peskin 2006; Sturrock et al. 2013; Winkelmann and Schutte 2016).

We will conclude this discussion with a summary of the main open questions we have encountered in our review of the stochastic reaction–diffusion field:

- Is there a simple way to calculate RDME hopping rates for oddly shaped subvolumes?
- Can we find RDME rate functions which converge to specific non-elementary functions (e.g. Michaelis–Menten) when diffusion is fast?
- How can we modify the RDME to incorporate hydrodynamic interactions?
- How can we unify the variety of different modifications to the diffusion coefficient to account for macromolecular crowding?
- How can we incorporate hydrodynamic interactions into BD simulations without incurring a huge computational cost?
- How effective is the RDME as a model of tissue dynamics?

Answering these questions will significantly advance the field and increase our ability to model larger and larger systems in reasonable time. If we were to select a problem which, to our mind, is the most significant obstacle for the field, it would be the issue of hydrodynamic interactions. Until we find a computationally efficient method to incorporate hydrodynamic effects into BD (or RDME) simulations, the fundamental physical validity of our simulations will potentially be in doubt.

Nonetheless, no matter which model is used, the RDME and BD are guaranteed to provide insight into the link between microscopic chemistry and macroscopic biology. As it gets computationally practical to simulate systems on increasingly large scales, the RDME and BD are likely to be the models which take us beyond the state of the art in whole-cell modelling (Karr et al. 2012) to provide the first single-molecule detailed whole-cell simulations and thereby allow us to visualise the macroscopic in terms of the microscopic for the first time.

References

Abramowitz M, Stegun IA (1964) Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol 55. Courier Corporation, Chelmsford

Agbanusi IC, Isaacson SA (2014) A comparison of bimolecular reaction models for stochastic reaction-diffusion systems. Bull Math Biol 76(4):922–946

Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. J Chem Phys 31(2):459–466

Alder BJ, Wainwright TE (1960) Studies in molecular dynamics. II. Behavior of a small number of elastic spheres. J Chem Phys 33(5):1439–1451

Ando T, Skolnick J (2010) Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proc Natl Acad Sci 107(43):18457–18462

Andrews SS, Bray D (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. Phys Biol 1(3):137

Andrews SS et al (2010) Detailed simulations of cell biology with Smoldyn 2.1. PLoS Comput Biol 6(3):e1000705

Andreychenko A et al (2017) Distribution approximations for the chemical master equation: comparison of the method of moments and the system size expansion. In: Modeling cellular systems. Springer, Cham, pp 39–66

Atkins P, de Paula J, Keeler J (2018) Atkins' physical chemistry. Oxford University Press, Oxford

Auger A, Chatelain P, Koumoutsakos P (2006) R-leaping: accelerating the stochastic simulation algorithm by reaction leaps. J Chem Phys 125(8):084103

Baker RE, Yates CA, Erban R (2010) From microscopic to macroscopic descriptions of cell migration on growing domains. Bull Math Biol 72(3):719–762

Ball K et al (2006) Asymptotic analysis of multiscale approximations to reaction networks. Ann Appl Probab 16(4):1925–1961

Banks DS, Fradin C (2005) Anomalous diffusion of proteins due to molecular crowding. Biophys J 89(5):2960–2971

Baras F, Mansour MM (1996) Reaction-diffusion master equation: a comparison with microscopic simulations. Phys Rev E 54(6):6139

Bayati B, Chatelain P, Koumoutsakos P (2011) Adaptive mesh refinement for stochastic reaction-diffusion processes. J Comput Phys 230(1):13–26

Bernstein D (2005) Simulating mesoscopic reaction-diffusion systems using the Gillespie algorithm. Phys Rev E 71(4):041103

Blum JJ et al (1989) Effect of cytoskeletal geometry on intracellular diffusion. Biophys J 56(5):995–1005

Borek B, Hasty J, Tsimring L (2016) Turing patterning using gene circuits with gas-induced degradation of quorum sensing molecules. PLoS ONE 11(5):e0153679

Bruna M, Chapman SJ (2012) Excluded-volume effects in the diffusion of hard spheres. Phys Rev E 85(1):011103

Buch I et al (2010) High-throughput all-atom molecular dynamics simulations using distributed computing. J Chem Inf Model 50(3):397–403

Butcher JC (2016) Numerical methods for ordinary differential equations. Wiley, New York

Cao Y, Gillespie DT, Petzold LR (2005) The slow-scale stochastic simulation algorithm. J Chem Phys 122(1):014116

Chapanian R et al (2014) Enhancement of biological reactions on cell surfaces via macromolecular crowding. Nat Commun 5:4683

Chen L et al (2010) Modeling biomolecular networks in cells: structures and dynamic. Springer, Berlin

Cianci C, Smith S, Grima R (2016) Molecular finite-size effects in stochastic models of equilibrium chemical systems. J Chem Phys 144(8):084101

Cianci C, Smith S, Grima R (2017) Capturing Brownian dynamics with an on-lattice model of hard-sphere diffusion. Phys Rev E 95(5):052118

Cichocki B, Hinsen K (1990) Dynamic computer simulation of concentrated hard sphere suspensions: I. Simulation technique and mean square displacement data. Phys A Stat Mech Appl 166(3):473–491

Collins FC, Kimball GE (1949) Diffusion-controlled reaction rates. J Colloid Sci 4(4):425–437

Darvey IG, Ninham BW, Staff PJ (1966) Stochastic models for secondorder chemical reaction kinetics. The equilibrium state. J Chem Phys 45(6):2145–2155

Dauty E, Verkman AS (2004) Molecular crowding reduces to a similar extent the diffusion of small solutes and macromolecules: measurement by fluorescence correlation spectroscopy. J Mol Recognit 17(5):441–447

Deutch JM, Felderhof BU (1973) Hydrodynamic effect in diffusioncontrolled reaction. J Chem Phys 59(4):1669–1671

Diambra L et al (2014) Cooperativity to increase Turing pattern space for synthetic biology. ACS Synth Biol 4(2):177–186

Ding F, Buldyrev SV, Dokholyan NV (2005) Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. Biophys J 88(1):147–155

Dirac PAM (1927) The quantum theory of the emission and absorption of radiation. Proc R Soc Lond A 114(767):243–265

Dix JA, Verkman AS (2008) Crowding effects on diffusion in solutions and cells. Annu Rev Biophys 37:247–263

Dobrzynski M et al (2007) Computational methods for diffusion-influenced biochemical reactions. Bioinformatics 23(15):1969–1977

Doi M (1976) Stochastic theory of diffusion-controlled reaction. J Phys A Math Gen 9(9):1479

Drawert B et al (2010) The diffusive finite state projection algorithm for efficient simulation of the stochastic reaction-diffusion master equation. J Chem Phys 132(7):074101

Drawert B, Engblom S, Hellander A (2012) URDME: a modular framework for stochastic simulation of reaction-transport processes in complex geometries. BMC Syst Biol 6(1):76

Durrant JD, McCammon JA (2011) BMC Biol. Molecular dynamics simulations and drug discovery 9(1):71

Einstein A (1956) Investigations on the theory of the Brownian Movement. Courier Corporation, Chelmsford

Elderfield D (1985) Field theories for kinetic growth models. J Phys A Math Gen 18(13):L773

Elf J, Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. Genome Res 13(11):2475–2484

Elf J, Ehrenberg M (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Syst Biol 1(2):230–236

Elf J, Doncic A, Ehrenberg M (2003) Mesoscopic reaction-diffusion in intracellular signaling. In: Fluctuations and noise in biological, biophysical, and biomedical systems, vol 5110. International Society for Optics and Photonics

Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26(10):597–604

Engblom S et al (2009) Simulation of stochastic reaction-diffusion processes on unstructured meshes. SIAM J Sci Comput 31(3):1774–1797

Erban R (2014) From molecular dynamics to Brownian dynamics. Proc R Soc A 470(2167):20140036

Erban R, Chapman J, Maini P (2007) A practical guide to stochastic simulations of reaction-diffusion processes. arXiv preprint arXiv:0704.1908

Erban R, Chapman SJ (2009) Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions. Phys Biol 6(4):046001

Erdi P, Janos T (1989) Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models. Manchester University Press, Manchester

Erdmann T, Howard M, ten Wolde PR (2009) Role of spatial averaging in the precision of gene expression patterns. Phys Rev Lett 103(25):258101

Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. J Chem Phys 69(4):1352–1360

Fanelli D, McKane AJ (2010) Diffusion in a crowded environment. Phys Rev E 82(2):021113

Fange D, Elf J (2006) Noise-induced min phenotypes in *E. coli*. PLoS Comput Biol 2(6):e80

Fange D et al (2010) Stochastic reaction-diffusion kinetics in the microscopic limit. Proc Natl Acad Sci 107(46):19820–19825

Fange D, Mahmutovic A, Elf J (2012) MesoRD 1.0: stochastic reaction-diffusion simulations in the microscopic limit. Bioinformatics 28(23):3155–3157

Felderhof BU (1978) Diffusion of interacting Brownian particles. J Phys A Math Gen 11(5):929

Fermi E (1950) Nuclear physics: a course given by Enrico Fermi at the University of Chicago. University of Chicago Press, Chicago

Flegg MB et al (2014) Analysis of the two-regime method on square meshes. SIAM J Sci Comput 36(3):B561–B588

Frenkel D, Smit B (2001) Understanding molecular simulation: from algorithms to applications, vol 1. Elsevier, New York

Fu J et al (2014) The time dependent propensity function for acceleration of spatial stochastic simulation of reaction-diffusion systems. J Comput Phys 274:524–549

Gadgil C, Lee CH, Othmer HG (2005) A stochastic analysis of first-order reaction networks. Bull Math Biol 67(5):901–946

Galanti M et al (2014) Diffusion of tagged particles in a crowded medium. EPL (Europhys Lett) 107(2):20006

Gardiner CW (1986) Handbook of stochastic methods for physics, chemistry and the natural sciences. Appl Opt 25:3145

Ghosh A, Leier A, Marquez-Lago TT (2015) The spatial chemical Langevin equation and reaction diffusion master equations: moments and qualitative solutions. Theor Biol Med Model 12(1):5

Gierer A, Meinhardt H (1972) A theory of biological pattern formation. Kybernetik 12(1):30–39

Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. J Phys Chem 81(25):2340–2361

Gillespie DT (1992) A rigorous derivation of the chemical master equation. Phys A Stat Mech Appl 188(1–3):404–425

Gillespie DT (2000) The chemical Langevin equation. J Chem Phys 113(1):297–306

Gillespie DT (2001) Approximate accelerated stochastic simulation of chemically reacting systems. J Chem Phys 115(4):1716–1733

Gillespie DT (2009a) A diffusional bimolecular propensity function. J Chem Phys 131(16):164109

Gillespie CS (2009b) Moment-closure approximations for mass-action models. IET Syst Biol 3(1):52–58

Gillespie DT, Seitaridou E (2012) Simple Brownian diffusion: an introduction to the standard theoretical models. Oxford University Press, Oxford

Gillespie DT, Hellander A, Petzold LR (2013) Perspective: stochastic algorithms for chemical kinetics. J Chem Phys 138(17):05B201

Gillespie DT, Petzold LR, Seitaridou E (2014) Validity conditions for stochastic chemical kinetics in diffusion-limited systems. J Chem Phys 140(5):02B604

Grima R (2012) A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. J Chem Phys 136(15):04B616

Grima R, Schnell S (2006) A systematic investigation of the rate laws valid in intracellular environments. Biophys Chem 124(1):1–10

Grima R, Schnell S (2008) Modelling reaction kinetics inside cells. Essays Biochem 45:41–56

Grima R, Schmidt DR, Newman TJ (2012) Steady-state fluctuations of a genetic feedback loop: an exact solution. J Chem Phys 137(3):035104

Gupta PB et al (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell 146(4):633–644

Hellander S, Hellander A, Petzold L (2012) Reaction-diffusion master equation in the microscopic limit. Phys Rev E 85(4):042901

Hellander S, Hellander A, Petzold L (2015) Reaction rates for mesoscopic reaction-diffusion kinetics. Phys Rev E 91(2):023312

Hepp B, Gupta A, Khammash M (2015) Adaptive hybrid simulations for multiscale stochastic reaction networks. J Chem Phys 142(3):034118

Hofling F, Franosch T (2013) Anomalous transport in the crowded world of biological cells. Rep Prog Phys 76(4):046602

Hu H, Yang W (2008) Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. Annu Rev Phys Chem 59:573–601

Isaacson SA (2008) Relationship between the reaction-diffusion master equation and particle tracking models. J Phys A Math Theor 41(6):065003

Isaacson SA (2009) The reaction-diffusion master equation as an asymptotic approximation of diffusion to a small target. SIAM J Appl Math 70(1):77–111

Isaacson SA (2013) A convergent reaction-diffusion master equation. J Chem Phys 139(5):054101

Isaacson SA, Peskin CS (2006) Incorporating diffusion in complex geometries into stochastic chemical kinetics simulations. SIAM J Sci Comput 28(1):47–74

Jahnke T, Huisinga W (2007) Solving the chemical master equation for monomolecular reaction systems analytically. J Math Biol 54(1):1–26

Kang H-W, Zheng L, Othmer HG (2012) A new method for choosing the computational cell in stochastic reaction-diffusion systems. J Math Biol 65(6–7):1017–1099

Karlin S (2014) A first course in stochastic processes. Academic Press, Cambridge

Karr JR et al (2012) A whole-cell computational model predicts phenotype from genotype. Cell 150(2):389–401

Klann M, Koeppl H (2012) Spatial simulations in systems biology: from molecules to cells. Int J Mol Sci 13(6):7798–7827

Klann MT, Lapin A, Reuss M (2009) Stochastic simulation of signal transduction: impact of the cellular architecture on diffusion. Biophys J 96(12):5122–5129

Klepeis JL et al (2009) Long-timescale molecular dynamics simulations of protein structure and function. Curr Opin Struct Biol 19(2):120–127

Kurtz TG (1971) Limit theorems for sequences of jump Markov processes. J Appl Probab 8(2):344–356

Kurtz TG (1972) The relationship between stochastic and deterministic models for chemical reactions. J Chem Phys 57(7):2976–2978

Lawson MJ, Petzold L, Hellander A (2015) Accuracy of the MichaelisMenten approximation when analysing effects of molecular noise. J R Soc Interface 12(106):20150054

L'Ecuyer P (2012) Random number generation. Handbook of computational statistics. Springer, Berlin, pp 35–71

Lengyel I, Epstein IR (1992) A chemical approach to designing Turing patterns in reaction–diffusion systems. Proc Natl Acad Sci 89(9):3977–3979

Lipkov J et al (2011) Analysis of Brownian dynamics simulations of reversible bimolecular reactions. SIAM J Appl Math 71(3):714–730

Lipkow K, Andrews SS, Bray D (2005) Simulated diffusion of phosphorylated CheY through the cytoplasm of *Escherichia coli*. J Bacteriol 187(1):45–53

Loeve M (1977) Elementary probability theory. Springer, New York, pp 1–52

Lotstedt P, Meinecke L (2015) Simulation of stochastic diffusion via first exit times. J Comput Phys 300:862–886

MacKerell AD Jr (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616

Marquez-Lago TT, Leier A, Burrage K (2012) Anomalous diffusion and multifractional Brownian motion: simulating molecular crowding and physical obstacles in systems biology. IET Syst Biol 6(4):134–142

McCammon JA, Gelin BR, Karplus M (1977) Dynamics of folded proteins. Nature 267(5612):585

McGuffee SR, Elcock AH (2010) Diffusion, crowding and protein stability in a dynamic molecular model of the bacterial cytoplasm. PLoS Comput Biol 6(3):e1000694

Meinecke L (2017) Multiscale modeling of diffusion in a crowded environment. Bull Math Biol 79(11):2672–2695

Meinecke L, Ltstedt P (2016) Stochastic diffusion processes on Cartesian meshes. J Comput Appl Math 294:1–11

Meinecke L et al (2016) Analysis and design of jump coefficients in discrete stochastic diffusion models. SIAM J Sci Comput 38(1):A55–A83

Metzler R, Klafter J (2000) The random walk's guide to anomalous diffusion: a fractional dynamics approach. Phys Rep 339(1):1–77

Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. J Chem Phys 124(4):044104

Murray JD (2001) Mathematical biology II: spatial models and biomedical applications. Springer, New York

Nakamasu A et al (2009) Interactions between zebrafish pigment cells responsible for the generation of Turing patterns. Proc Natl Acad Sci 106(21):8429–8434

Ohtsuki T, Okano K (1982) Diffusion coefficients of interacting Brownian particles. J Chem Phys 77(3):1443–1450

Opplestrup T et al (2006) First-passage Monte Carlo algorithm: diffusion without all the hops. Phys Rev Lett 97(23):230602

Pastor RW, Brooks BR, Szabo A (1988) An analysis of the accuracy of Langevin and molecular dynamics algorithms. Mol Phys 65(6):1409–1419

Purcell EM (2014) Life at low Reynolds number. In: Physics and our world: reissue of the proceedings of a symposium in honor of Victor F Weisskopf

Raspopovic J et al (2014) Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. Science 345(6196):566–570

Redner S (2001) A guide to first-passage processes. Cambridge University Press, Cambridge

Reverey JF et al (2015) Superdiffusion dominates intracellular particle motion in the supercrowded cytoplasm of pathogenic *Acanthamoeba castellanii*. Sci Rep 5:11690

Rickard D (1997) Kinetics of pyrite formation by the H2S oxidation of iron (II) monosulfide in aqueous solutions between 25 and 125 C: the rate equation. Geochim Cosmochim Acta 61(1):115–134

Ridgway D et al (2008) Coarse-grained molecular simulation of diffusion and reaction kinetics in a crowded virtual cytoplasm. Biophys J 94(10):3748–3759

Riniker S, van Gunsteren WF (2011) A simple, efficient polarizable coarse-grained water model for molecular dynamics simulations. J Chem Phys 134(8):084110

Roberts E, Stone JE, LutheySchulten Z (2013) Lattice microbes: highperformance stochastic simulation method for the reaction-diffusion master equation. J Comput Chem 34(3):245–255

Rotne J, Prager S (1969) Variational treatment of hydrodynamic interaction in polymers. J Chem Phys 50(11):4831–4837

Salis H, Kaznessis Y (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J Chem Phys 122(5):054103

Saxton MJ (1994) Anomalous diffusion due to obstacles: a Monte Carlo study. Biophys J 66(2):394–401

Schnell S, Mendoza C (1997) Closed form solution for time-dependent enzyme kinetics. J Theor Biol 187(2):207–212

Schnoerr D, Sanguinetti G, Grima R (2014a) Validity conditions for moment closure approximations in stochastic chemical kinetics. J Chem Phys 141(8):08B616

Schnoerr D, Sanguinetti G, Grima R (2014b) The complex chemical Langevin equation. J Chem Phys 141(2):07B606

Schnoerr D, Sanguinetti G, Grima R (2015) Comparison of different moment-closure approximations for stochastic chemical kinetics. J Chem Phys 143(18):11B610

Schnoerr D, Sanguinetti G, Grima R (2017) Approximation and inference methods for stochastic biochemical kineticsa tutorial review. J Phys A Math Theor 50(9):093001

Scholes NS, Isalan M (2017) A three-step framework for programming pattern formation. Curr Opin Chem Biol 40:1–7

Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci 105(45):17256–17261

Sheth R et al (2012) Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. Science 338(6113):1476–1480

Singh A, Hespanha JP (2011) Approximate moment dynamics for chemically reacting systems. IEEE Trans Autom Control 56(2):414–418

Smith S, Dalchau N (2018a) Model reduction enables turing instability analysis of large reaction-diffusion models. J R Soc Interface 15(140):20170805

Smith S, Dalchau N (2018b) Beyond activator-inhibitor networks: the generalised Turing mechanism. arXiv preprint arXiv:1803.07886

Smith S, Grima R (2016) Breakdown of the reaction-diffusion master equation with nonelementary rates. Phys Rev E 93(5):052135

Smith S, Grima R (2017a) Fast simulation of Brownian dynamics in a crowded environment. J Chem Phys 146(2):024105

Smith S, Grima R (2017b) Model reduction for stochastic reaction systems. In: Stochastic processes, multiscale modeling, and numerical methods for computational cellular biology, Springer, Cham, pp 143–158

Smith S, Grima R (2018) Single-cell variability in multicellular organisms. Nat Commun 9(1):345

Smith AV, Hall CK (2001) Protein refolding Versus aggregation: computer simulations on an intermediate-resolution protein model1. J Mol Biol 312(1):187–202

Smith S, Shahrezaei V (2015) General transient solution of the one-step master equation in one dimension. Phys Rev E 91(6):062119

Smith S, Cianci C, Grima R (2015) Model reduction for stochastic chemical systems with abundant species. J Chem Phys 143(21):12B615

Smith S, Cianci C, Grima R (2016) Analytical approximations for spatial stochastic gene expression in single cells and tissues. J R Soc Interface 13(118):20151051

Smith S, Cianci C, Grima R (2017) Macromolecular crowding directs the motion of small molecules inside cells. J R Soc Interface 14(131):20170047

Smoluchowski MV (1917) An experiment on mathematical theorization of coagulation kinetics of the colloidal solutions. Z Phys Chem 92:129–168

Smoluchowski M (1924) Letters of Marian. Smoluchowski 1:479

Stundzia AB, Lumsden CJ (1996) Stochastic simulation of coupled reaction-diffusion processes. J Comput Phys 127(1):196–207

Sturrock M et al (2013) Spatial stochastic modelling of the Hes1 gene regulatory network: intrinsic noise can explain heterogeneity in embryonic stem cell differentiation. J R Soc Interface 10(80):20120988

Szekely T, Burrage K (2014) Stochastic simulation in systems biology. Comput Struct Biotechnol J 12(20):14–25

Takahashi K, Tanase-Nicola S, ten Wolde PR (2010) Spatio-temporal correlations can drastically change the response of a MAPK pathway. Proc Natl Acad Sci 107(6):2473–2478

Tan C et al (2013) Molecular crowding shapes gene expression in synthetic cellular nanosystems. Nat Nanotechnol 8(8):602

Taylor PR, Baker RE, Yates CA (2014) Deriving appropriate boundary conditions, and accelerating position-jump simulations, of diffusion using non-local jumping. Phys Biol 12(1):016006

Teramoto E, Shigesada N (1967) Theory of bimolecular reaction processes in liquids. Prog Theor Phys 37(1):29–51

Thomas P, Grima R (2015) Approximate probability distributions of the master equation. Phys Rev E 92(1):012120

Torquato S, Stillinger FH (2010) Jammed hard-particle packings: from Kepler to Bernal and beyond. Rev Mod Phys 82(3):2633

Tozzini V (2005) Coarse-grained models for proteins. Curr Opin Struct Biol 15(2):144–150

Turing AM (1952) The chemical basis of morphogenesis. Philos Trans R Soc Lond Ser B Biol Sci 237(641):37–72

Twycross J et al (2010) Stochastic and deterministic multiscale models for systems biology: an auxin-transport case study. BMC Syst Biol 4(1):34

Van Kampen NG (1976) The equilibrium distribution of a chemical mixture. Phys Lett A 59(5):333–334

Van Kampen NG (1992) Stochastic processes in physics and chemistry, vol 1. Elsevier, New York

van Zon JS, ten Wolde PR (2005) Greens-functionreaction dynamics: a particle-based approach for simulating biochemical networks in time and space. J Chem Phys 123(23):234910

Wang H, Junghans C, Kremer K (2009) Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? Eur Phys J E 28(2):221–229

Watanabe M, Kondo S (2015) Is pigment patterning in fish skin determined by the Turing mechanism? Trends Genet 31(2):88–96

Weiss M et al (2004) Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. Biophys J 87(5):3518–3524

Weissberg HL (1963) Effective diffusion coefficient in porous media. J Appl Phys 34(9):2636–2639

Winkelmann S, Schutte C (2016) The spatiotemporal master equation: approximation of reaction-diffusion dynamics via Markov state modeling. J Chem Phys 145(21):214107

Wojcieszyn JW, Schlegel RA, Jacobson KA (1982) Measurements of the diffusion of macromolecules injected into the cytoplasm of living cells. In: Cold Spring Harbor symposia on quantitative biology, vol 46, Cold Spring Harbor Laboratory Press

Yamakawa H (1971) Modern theory of polymer solutions. Harper & Row, New York

Zimmerman SB, Minton AP (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. Ann Rev Biophys Biomol Struct 22(1):27–65

Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222(3):599–620