Chemistry at the Nanoscale

When Every Reaction is a Discrete Event

Ashwin B R Kumar and Ram Ramaswamy

Traditionally the kinetics of a chemical reaction has been studied as a set of coupled ordinary differential equations. The law of mass action, a tried and tested principle for reactions involving macroscopic quantities of reactants, gives rise to deterministic equations in which the variables are species concentrations. In recent years, though, as smaller and smaller systems – such as an individual biological cell, say – can be studied quantitatively, the importance of molecular discreteness in chemical reactions has increasingly been realized. This is particularly true when the system is far from the 'thermodynamic limit' when the numbers of all reacting molecular species involved are several orders of magnitude smaller than Avogadro's number. In such situations, each reaction has to be treated as a probabilistic 'event' that occurs by chance when the appropriate reactants collide. Explicitly accounting for such processes has led to the development of sophisticated statistical methods for simulation of chemical reactions, particularly those occurring at the cellular and sub-cellular level. In this article, we describe this approach, the so-called stochastic simulation algorithm, and discuss applications to study the dynamics of model regulatory networks.

Ashwin B R Kumar is a graduate student in the Biophysics, Structural and Computational Biology Program, School of Medicine and Dentistry, University of Rochester, New York. His research interests are in studying the structure, function, interaction and the chemistry of biomolecules as well as their assemblies and networks.

Ram Ramaswamy teaches in the Schools of Physical Sciences (SPS) and Computational and Integrative Sciences (SCIS) at Jawaharlal Nehru University, New Delhi. His research focuses on nonlinear science as well as computational systems biology.

1. Introduction

The subject of *chemical kinetics* essentially relates to the study of the progression of chemical reactions and the rates of transformation of chemical species from reactants to products [1, 2]. The discipline traces its origins to 1850 when the rate of inversion of sucrose into glucose and fructose was modelled by the German chemist Ludwig Wilhelmy who used an ordinary differential

Keywords
Chemical kinetics, stochastic chemical kinetics, Gillespie's stochastic simulation algorithm, synthetic gene oscillators.

equation (ODE) to describe the process mathematically. Shortly thereafter, Guldberg and Waage proposed the law of mass action, after which the subject developed rapidly, with important contributions coming from van't Hoff, Arrhenius and Ostwald among others. The importance of understanding how chemical reactions occurred, and what factors affected their rates or efficiency has always been a matter of great practical importance. The major advances of the late 1800s were quickly recognised with Nobel Prizes, the very first in 1901 going to van't Hoff for his work on the development of chemical kinetics, chemical dynamics, and on the concept of osmotic pressure. Arrhenius' derivation of the eponymous rate equation led to his Prize in 1903, and in 1909 to Ostwald, for several contributions to physical chemistry that included the principle of independence of chemical reactions, and the study of catalysis and chemical kinetics.

The classical approach to study the kinetics of a chemical reaction is to use one or more ODEs to analyse the time evolution of concentrations of the various chemical species involved. This approach, although recognising that chemical reactions are binary events on the molecular scale, encapsulates the details of all interactions into effective rates that apply when the chemical species are abundant. The concept of the *order* of a reaction and the dependence of the rate on the stoichiometry of the reaction implies that the method for modelling chemical reactions is in terms of continuous variables that obey deterministic equations of motion.

The classical approach to study the kinetics of a chemical reaction is to use one or more ODEs to analyse the time evolution of concentrations of the various chemical species involved. This approach, although recognising that chemical reactions are binary events on the molecular scale, encapsulates the details of all interactions into effective rates that apply when the chemical species are abundant.

However, this approach is inadequate when the populations of the chemical species are very low. For instance, in a typical biological cell, the total number of molecules ranges from 10^9 (for bacteria) to 10^{13} (for a eukaryotic cell) or so. The number of molecules of biological interest is much smaller than this, ranging from one (for DNA) to a few hundred RNA to a few thousand proteins. Although the effective concentrations are in the nano- or sub-nanomolar range, advances in experimental techniques to study single molecules have made it possible to investigate the dynamics (and hence also the kinetics) at this scale. In such systems, a classical kinetics approach cannot accurately predict the

behaviour; most of the assumptions underlying the law of mass action and kinetic theory are not valid.

At the microscopic level, chemical reactions occur when two molecules come close enough, in the proper relative orientations, and with the appropriate relative energies and angular momentum to interact, namely, either to make new chemical bonds or to break existing ones and sometimes both. The reaction mechanism for any given reaction is a way of interpreting this statement on a microscopic basis, while the macroscopic rate laws are a means of rationalising experimental results. Clearly, both these descriptions mask a wealth of detail since at the microscopic level molecules scatter off each other with specific energy and angular momentum, and the probability of reaction is related to the scattering cross section. Such quantities cannot be computed with quantitative accuracy for any but the most elementary reactions: the calculations are straightforward but tedious, and essential inputs such as interaction potentials are not known to sufficient accuracy.

At the microscopic level, chemical reactions occur when two molecules come close enough, in the proper relative orientations, and with the appropriate relative energies and angular momentum to interact.

An intermediate – mesoscopic – approach recognises the fact that when the number of molecules is not very large, a probabilistic approach can be used. Thus, the collision of two reactant species can be thought of as a random event that depends on the number of each of them and the volume of the container. Each such collision leads to a reaction with some probability, and each reaction changes the number of each species in a discrete fashion. When there are several possible reactions, depending on the number and types of species, these are taken to be concurrent random processes; the randomness and discreteness become very significant when the numbers are small and far from the so-called thermodynamic limit. The stochastic nature of chemical reactions in such situations becomes significant since noise plays a major role in the evolution and response of such systems.

Stochastic approaches to the study of chemical reactions can be traced back to the early work of Kramers [3] and Delbrück [4], who modelled the stochastic dynamics of an autocatalytic reaction, followed by numerous other studies that have been summa-

The stochastic simulation algorithm (SSA) proposed by Gillespie is widely used to simulate the dynamics of chemical and biochemical reactions in nanoscale environments.

rized in the very influential review article by McQuarrie in 1967 [5]. The practical implementation of this methodology was enunciated in a very lucid article by Gillespie [6] a decade later. Today, his formulation of the stochastic simulation algorithm (SSA) is widely used to simulate the dynamics of chemical and biochemical reactions in nanoscale environments. In addition to the already noted applications to cellular and subcellular processes, similar considerations will also apply to chemical reactions occurring on (say) dust grains in interstellar space where temperatures and densities are very low, and the probability of molecular encounters are consequently very small.

2. Stochastic Simulation Algorithm

The law of mass action that is taught early in all chemical kinetics courses states that the rate of a chemical reaction is proportional to the product of the concentrations of the reacting substances, with each concentration raised a power that is the stoichiometric coefficient in the corresponding chemical equation. The fact that this is only true in limited cases becomes clear in even simple examples such as the reaction between hydrogen and iodine chloride, with the formation of iodine and hydrogen chloride [2],

$$H_2 + 2ICl \rightarrow I_2 + 2HCl$$

for which one would write (incorrectly, as it turns out) the rate as proportional to:

$$[H_2][ICl]^2$$

(where [X] represents the concentration of X), which is a third-order reaction. Experimentally though, the reaction is seen to be of second-order, consistent with the microscopic description comprising two consecutive reactions,

$$H_2 + ICl \rightarrow HI + HCl \quad \text{and} \quad HI + ICl \rightarrow I_2 + 2HCl$$

The first reaction is slow while the second is much more rapid. Thus, the rate is actually proportional to $[H_2][ICl]$, making it effectively a reaction of second-order.

The Gillespie approach is to treat such consecutive or concurrent reactions within a general framework. Say there are N chemical species, X_1, X_2, X_3, ..., X_N that can participate in M distinct reactions, each of which has the form:

$$\text{Reactants} \rightarrow \text{Products}$$

The 'Products' on the right-hand side can be one or more of the molecules being considered here, namely from within the set of the X_i's, in which case these can subsequently participate in one or the other of the M reactions that are possible. The products can also be other molecules that do not participate in the reactions being considered, in which case these are denoted by $\emptyset$. Similarly, the reactants can be from the set of the X_i's or can include other molecules, as will be made clear in the examples discussed below.

The configuration at any instant of time, namely the numbers of the different molecules that are present is denoted by the integers $n_1, n_2, \ldots, n_N$. Each of the reactions is treated as a Poisson random process that occurs at a specific rate that depends on the propensity of the reaction. The propensity for each of the reactions depends on the configuration (see *Box* 1 for details of how these are computed). The essential component of the Gillespie algorithm is a recipe for determining *which* of the M different reactions will actually occur, and more importantly, *when* it will occur. At each step, therefore, one determines the time of the next reaction by generating an exponentially distributed random number which is based on the overall rate of *any* reaction occurring, given the specific configuration and the resulting propensities.

Therefore, given a configuration at a point in time, one can find out when the next reaction will occur as well as which reaction it will be. One simultaneously advances the internal time and changes the configuration, depending on the stoichiometry of the reaction that occurs. It is important to note that the direct version of the SSA is statistically *exact*. Being simple to understand, it is also easy to implement although it can become computationally slow and expensive when the complexity of the system under consideration increases. At present, there are a number of

The essential component of the Gillespie algorithm is a recipe for determining *which* of the M different reactions will actually occur, and more importantly, *when* it will occur.

For a proper 'systems level' understanding of biological processes, a computational approach is useful owing to the inherent complexity of the various interacting processes that are involved in even the simplest cases.

improvements to the basic algorithm and alternative approaches that have made stochastic simulation methods considerably faster [7].

3. Modelling Stochastic Dynamics

Current experimental techniques allow for the measurement of the rates of many elementary reactions. For a proper 'systems level' understanding of biological processes, however, a computational approach is useful owing to the inherent complexity of the various interacting processes that are involved in even the simplest cases. In conjunction with experiments, computational modelling can, therefore, provide some insight into the system dynamics.

Implementation of SSA is best discussed in the context of simple model systems for which the dynamics can be analysed in detail. One such set of coupled autocatalytic reactions that was proposed by Schlögl [8] consists of four reactions R_{1-4} involving three species denoted as A, B, and X:

$$\begin{aligned} R_1 &: \quad A + 2X \xrightarrow{c_1} 3X \\ R_2 &: \quad 3X \xrightarrow{c_2} A + 2X \\ R_3 &: \quad B \xrightarrow{c_3} X \\ R_4 &: \quad X \xrightarrow{c_4} B \end{aligned}$$

where the rates of the equations are indicated by the c_i's. The populations of A and B are taken to remain constant and much larger than that of the species of interest, namely X. For this dynamics, it is straightforward to obtain the kinetic equation:

$$\frac{dx}{dt} = c_1 a x^2 - c_2 x^3 + c_3 b - c_4 x,$$

where a, b, and x are the concentrations of A, B, and X respectively. Although nonlinear, the above equation can be easily solved for specific values of the rate constants c_i's and a and b; typical results are shown in *Figure* 1.

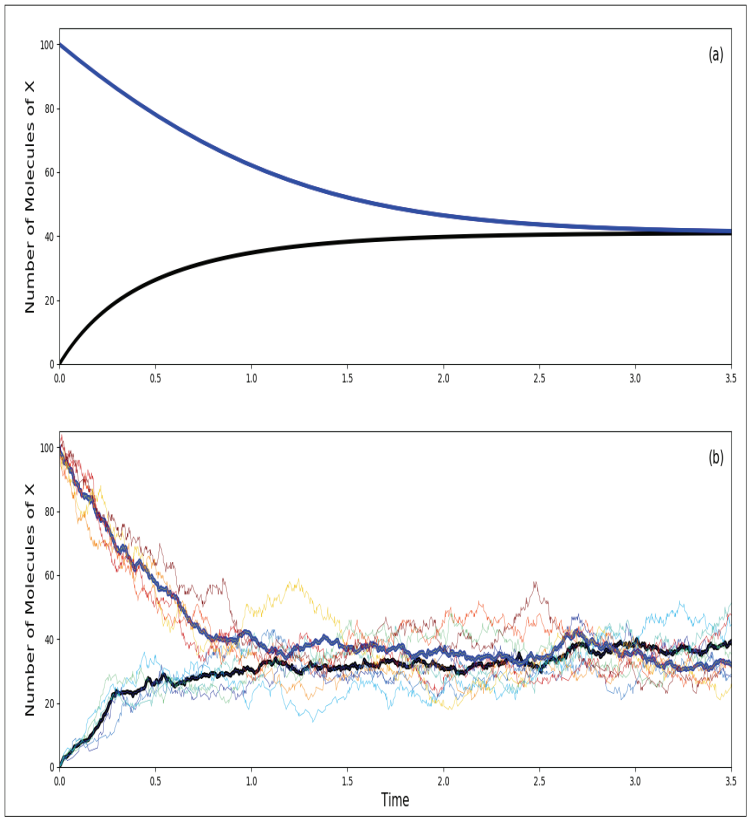

Figure 1. Panel **(a)** shows the temporal behaviour of X obtained by solving the kinetic equation for parameter values: $a = b = 10^5$, and $c_1 = 3{\cdot}10^{-7}$, $c_2 = 10^{-4}$, $c_3 = 10^{-3}$, and $c_4 = 3.5$ for two different values of the initial concentration of X which in steady-state reaches a value of 40. Panel **(b)** shows the results from a stochastic simulation. The bold lines in (b) are averages of several independent runs carried out with similar initial conditions as in (a). As can be seen, there is clearly a similarity in the average stochastic behaviour and the results of the deterministic simulation, the agreement getting better as the system size is increased.

The key contrast between stochastic and deterministic modelling of a reaction system can be seen in this example. As we can see from the simulation of the Schlögl reactions shown in *Figure* 1(b), each stochastic run gives a different trajectory for the time evolution of the reaction (each thin line represents a specific run), whereas in the deterministic modelling approach every simulation will give the same trajectory for the specified initial conditions. One, therefore, takes an average of several stochastic simulations (the bold lines in *Figure* 1(b)) in order to make a proper comparison with the kinetic simulations in *Figure* 1(a).

3.1 *Dynamics of Biological Systems*

Periodic variations in the concentrations of specific biomolecules play an extremely important role in regulation of the biological processes they are invloved in, a feature that has been studied extensively in naturally occurring systems as well as in synthetic oscillators.

This approach to chemical kinetics is particularly suited to the study of biological systems since, as pointed out earlier, many of the reactant species occur in small numbers within cells. Indeed, one of the earliest applications was to the kinetics of the genetic switch in the λ-phage system [9]. By now, there are numerous experiments that have directly shown the influence of stochasticity in biochemical reactions in both prokaryotic and eukaryotic cells.

Here we focus on oscillatory behaviour. The maintenance of rhythms is extremely important in biology – many biological 'clocks' play a crucial role in the functioning of organisms. Oscillations in biology are known to range from timescales of milliseconds as in neuronal processes, to seconds as in calcium oscillations or cardiac rhythms, to circadian clocks that have a periodicity of about one day or 24 hours. Longer periodicities are also known, such as ovarian cycles that last a month, and ecological cycles that have timescales of years. All such rhythms are dominated by fluctuations. In fact, to recognise and celebrate the importance of the research on biological clocks, the Nobel Prize in Physiology or Medicine for the year 2017 was awarded to Hall, Rosbash and Young for their research on molecular mechanisms of the circadian rhythm[1].

[1] For interesting information on 'circadian rhythms', see Series Article by K M Vaze, V K Sharma and K L Nikhil, *Resonance*, Vol.18, No.7, 9 and 11, 2013; Vol.19, No.2, 2014.

The area of systems biology, which aims to study the dynamics and behaviour of a network of biological reactions is currently of great interest. In most biological systems, there are several reacting biochemical species, and an interesting aspect of the dynamics in such networks is sustained oscillation in the concentration of key molecules. Such periodic variations play an extremely important role in regulation, a feature that has been studied extensively in naturally occurring systems as well as in synthetic oscillators.

An early example of a synthetic gene oscillator model with oscillations was proposed by Goodwin [10] in which a gene produces a protein that represses its own expression as shown in *Figure 2*.

This model consists of the following six reactions:

$$
\begin{aligned}
R_1 : &\quad D \xrightarrow{c_1} D + M \\
R_2 : &\quad M \xrightarrow{c_2} M + P_1 \\
R_3 : &\quad P_1 \xrightarrow{c_3} P_2 \\
R_4 : &\quad M \xrightarrow{c_4} \emptyset \\
R_5 : &\quad P_1 \xrightarrow{c_5} \emptyset \\
R_6 : &\quad P_2 \xrightarrow{c_6} \emptyset
\end{aligned}
$$

that involve four molecular species. D and M correspond respectively to the promoter region on the DNA and the messenger RNA. P_1 is the protein product from M and P_2 is its transcriptional repressor form. The model has been used extensively for studying the dynamics of enzyme catalysis, transcriptional gene regulation, and multi-site protein phosphorylation processes.

The kinetic equations derived by Goodwin [10, 11] are:

$$
\begin{aligned}
\frac{dM}{dt} &= \frac{1}{1+P_2^n} - \alpha M\,, \\
\frac{dP_1}{dt} &= M - \beta P_1\,, \\
\frac{dP_2}{dt} &= P_1 - \gamma P_2\,,
\end{aligned}
$$

where M, P_1, and P_2 are the concentrations of M, P_1, and P_2 respectively. The negative feedback due to the inhibition caused by P_2 on the production of M is described by a Hill function, with

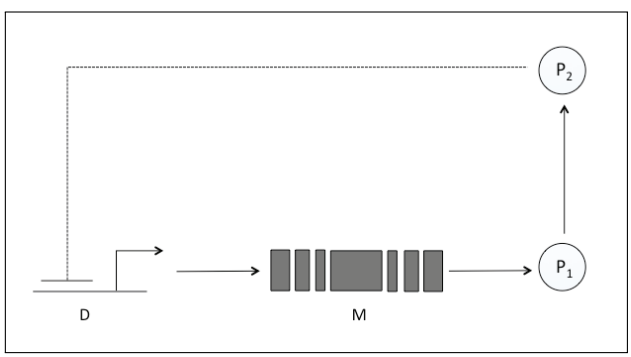

Figure 2. Biochemical network for the Goodwin model. D denotes the promoter region on the DNA, M denotes the mRNA, P_1 is the protein product, while P_2 is the transcriptional repressor form of P_1.

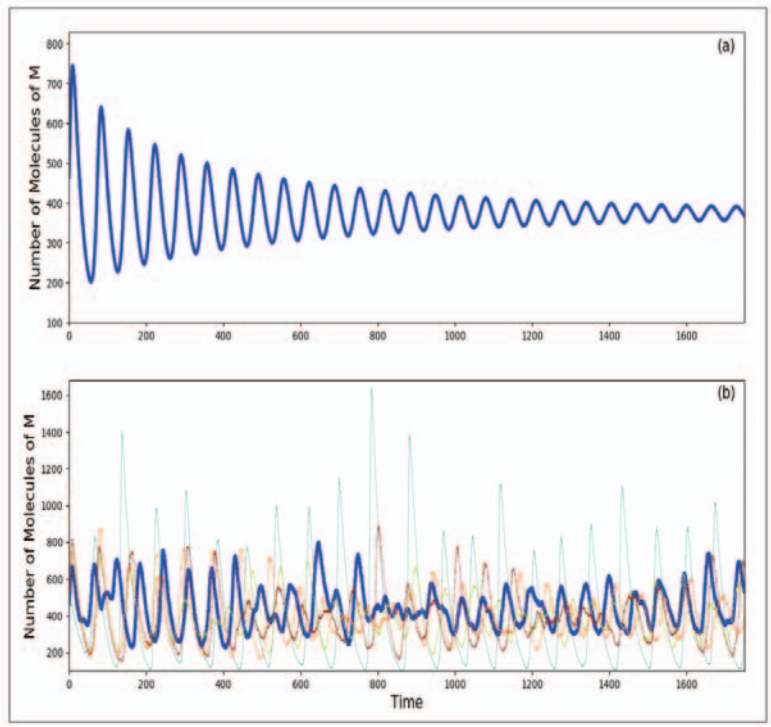

Figure 3. Dynamics of M in the Goodwin model for n=10. The other parameters have the values $c_2 = c_3 = 0.075$ min, $c_4 = c_5 = c_6 = 0.0375$ min^{-1} and $s = 300$. The deterministic dynamics is shown in **(a)** above and the corresponding stochastic dynamics is shown in **(b)** below. While the deterministic dynamics at these values of the parameters clearly leads to damped oscillations in the stochastic version, the motion is not as clearly damped. Such a difference between deterministic and stochastic simulations highlights the need for stochastic simulations when the systems are small.

the coefficient n being treated as a parameter. It is known [11] that below $n = 8$ the dynamics is damped, while for larger n there can be limit cycle oscillations for suitable values of other parameters.

Setting up the SSA for this system can be done in a straightforward manner, treating the negative inhibition as a modification of the basic rate c_1. When the system size s is included in the formalism it can be shown that this changes the effective values of the propensity as $c_1 = 3sK^n/4(K^n + P_2^n)$ molecules min^{-1}, where $K \equiv s$ molecules.

As can be seen in *Figures* 3 and 4, there are significant similarities as well as differences between the deterministic dynamics and the stochastic dynamics even when describing the same system, emphasising the need to use the appropriate formalism depending on the situation being modelled.

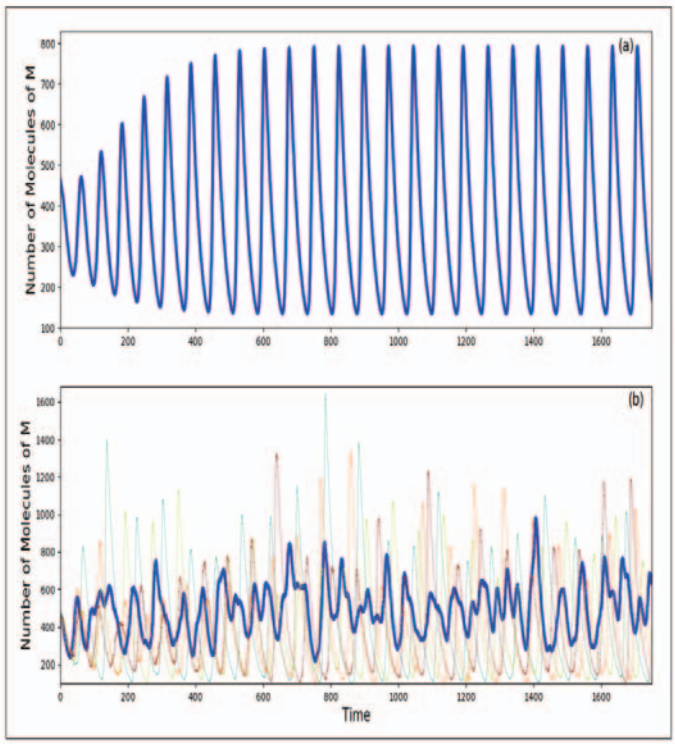

Figure 4. Dynamics of M in the Goodwin model for n=15; other parameters as in *Figure* 3. The deterministic dynamics is shown in **(a)** above and the corresponding stochastic dynamics is shown in **(b)** below. The deterministic dynamics at these values of the parameters clearly leads to a limit cycle with sustained oscillations. In the stochastic version, however, the dynamics while oscillatory, is not strictly periodic. This difference reduces as the volume increases.

Another simple model of biological oscillations was proposed by Vilar, Kueh, Barkai, and Leibler [12] (VKBL). The genetic network of this model, shown diagrammatically in *Figure* 5 is somewhat more complex than the simple feedback loop proposed by Goodwin and consists of two genes, an activator A and a repressor R in a negative feedback loop. The activator A promotes its own transcription as well as that of R by binding to the corresponding promoters. The repressor R acts negatively by sequestering A through the formation of an activator-repressor complex. This combination of positive and negative feedbacks yield sustained oscillations.

There are sixteen different processes that need to be considered in the VKBL network, and these give the following 'chemical'

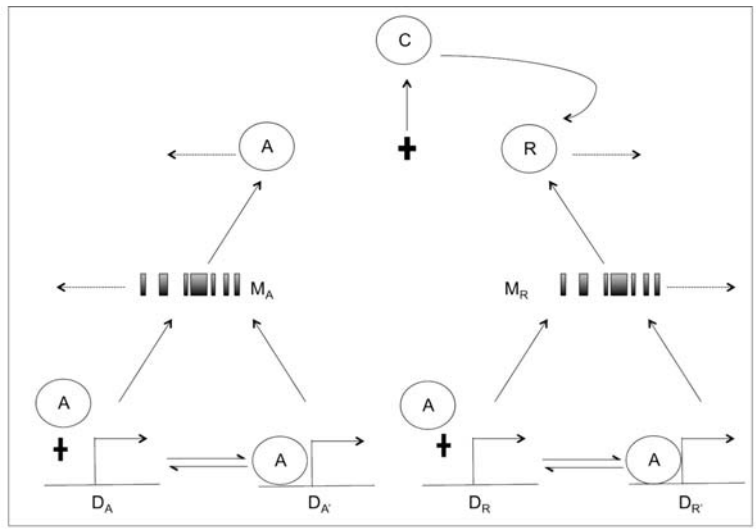

Figure 5. Biochemical network of the VKBL model [12]. D_A and D_R denotes the promoter regions on the DNA without the activator A bound for gene A or R; $D_{A'}$ and $D_{R'}$ denotes the promoter regions on the DNA with the activator A bound for gene A or R; M_A and M_R denotes the mRNA for A and R, while C represents the complex formed by A and R.

reactions:

$$
\begin{aligned}
&R_1: \quad D_A + A \xrightarrow{c_1} D_{A'} && R_9: A + R \xrightarrow{c_9} C\\
&R_2: \quad D_{A'} \xrightarrow{c_2} D_A + A && R_{10}: A \xrightarrow{c_{10}} \emptyset\\
&R_3: \quad D_R + A \xrightarrow{c_3} D_{R'} && R_{11}: D_{R'} \xrightarrow{c_{11}} D'_R + M_R\\
&R_4: \quad D_{R'} \xrightarrow{c_4} D_R + A && R_{12}: D_R \xrightarrow{c_{12}} D_R + M_R\\
&R_5: \quad D'_A \xrightarrow{c_5} D_A + M_A && R_{13}: M_R \xrightarrow{c_{13}} \emptyset\\
&R_6: \quad D_A \xrightarrow{c_6} D_A + M_A && R_{14}: M_R \xrightarrow{c_{14}} M_R + R\\
&R_7: \quad M_A \xrightarrow{c_7} \emptyset && R_{15}: C \xrightarrow{c_{15}} R\\
&R_8: \quad M_A \xrightarrow{c_8} M_A + A && R_{16}: R \xrightarrow{c_{16}} \emptyset
\end{aligned}
$$

Application of the stochastic simulation algorithm is fairly straightforward, given the above equations. The temporal variations in R and A obtained through stochastic simulations are shown in *Figure* 6. The reaction rates (see the caption of *Figure* 6 for their values) have been measured or estimated and putting all these together, one can clearly see that the concentrations of R and A vary in an oscillatory manner, with a period of about 24 hours.

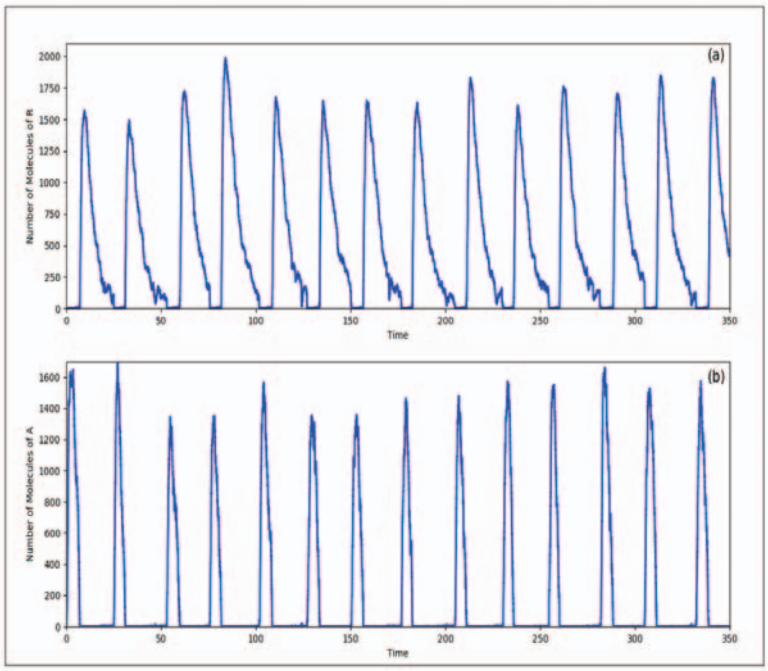

Figure 6. Results from the stochastic simulations of the VKBL model showing the variation of repressor in **(a)** and protein A in **(b)** as a function of time (in hours). The circadian nature of the oscillations is evident. In these simulations we have used the parameter values c_1 = 1 mol^{-1} h^{-1}; c_2 = 50 h^{-1}; c_3 = 1 $mol^{-1}h^{-1}$; c_4 = 100 h^{-1}; c_5 = 500 h^{-1}; c_6 = 50 h^{-1}; c_7 = 10 h^{-1}; c_8 = 50 h^{-1}; c_9 = 2 mol^{-1} h^{-1}; c_{10} = 1 h^{-1}; c_{11} = 50 h^{-1}; c_{12} = 0.01 h^{-1}; c_{13} = 0.5 h^{-1}; c_{14} = 5 h^{-1}; c_{15} = 1 h^{-1}; c_{16} = 0.2 h^{-1}. Some of these reaction constants have been adjusted so as to give the near 24 hour periodicity. The initial conditions are D_A = D_R = 1 mol, $D_{A'}$ = $D_{R'}$ = M_A =M_R = A=R=C=0.

4. Summary

Realistic simulations of biological systems have become possible since many of the basic rates – the c_μ's for many elementary processes – can now be measured, making the overall system amenable to modelling. This methodology has greatly enhanced our understanding of the existing genetic networks in biological cells and also helped in the design of new 'synthetic' regulatory modules that can be used to engineer novel biological and dynamical behaviour. Such modelling is an essential component of the *systems approach* that can be used to address complex biological issues. Some problems that have been explored through this technique include a study of the effect of miRNA on existing gene oscillators, the coupling of ensembles of genetic oscillators, the dynamics of regulatory modules, quorum sensing, natural and designed biological switches, and so on [13].

When dealing with nanoscale systems, a mesoscopic approach is often essential, in addition to being practical. The number of molecules of interest is much smaller than Avogadro's number but still too large to be treated accurately through atomistic simulations. Modelling processes at this scale require stochastic methods. This has also helped to clarify the effects of noise and fluctuations on the dynamics, both at the level of isolated systems as well as at that of populations. In addition, as can be appreciated, other effects that are crucial in realistic modelling can be incorporated fairly easily [14]. Diffusion can, for instance, be accounted for by the introduction of spatial degrees of freedom, and increasing the number of 'reactions' to allow for species to move and to diffuse, while spatial heterogeneity can be included by having differential site-dependent diffusion rates. In order to account for time delay, the algorithms are somewhat more complex but a variety of procedures, both exact and approximate, have been worked out [14].

Although we have mainly discussed biological examples, it should be noted that stochastic simulations have been also used to study the dynamics of other systems wherein the numbers of participating entities is not very large and/or when the natural spatial dimensions are nano-scalar. Both these limits are realized in a variety of situations that range from say the case of modelling epidemics and similar processes in population dynamics to the simulation and study of reactions in confined media or on surfaces.

Box 1. Outline of the Stochastic Simulation Algorithm

Gillespie [6] developed the SSA as follows. For each reaction R_μ it is necessary to determine the probability that it can occur in the time interval $(t, t + dt)$, given the configuration at time t. This probability, denoted $a_\mu dt$, is computed from the stoichiometry of the reaction as well as the reaction constant c_μ:

$$a_\mu dt = h_\mu c_\mu dt\,,$$

where h_μ depends on the stoichiometry and the configuration. Thus, if the reaction is of the type $X_i \rightarrow$ products, then $h_\mu = n_i$, namely the number of molecules of X_i available. If the reaction is of the form $X_i + X_j \rightarrow$ products, then clearly $h_\mu = n_i n_j$, and if it is of the type $2X_i \rightarrow$ products, then $h_\mu = n_i(n_i - 1)/2$, and so on. All kinetic and thermodynamic factors are subsumed in c_μ.

Given the instantaneous configuration and the reaction constants (which must be determined either via experiments or from other calculations), the relative probability of each of the reaction can be computed as $a_1, a_2, \ldots, a_M$. Denote by $P_0(\tau)$ the probability that no reaction occurs at time τ. Clearly, the probability that there is no reaction up to time $\tau + d\tau$ is given by:

$$P_0(\tau + d\tau) = P_0(\tau)\Big\{1 - \sum_{\mu=1}^{M} a_\mu d\tau\Big\}\,,$$

that is the probability that there was no reaction at time τ and further that no event occurs in the time interval $d\tau$. This is easily solved to give:

$$P_0(\tau) = \exp[-a_0\tau]\,,$$

where a_0 is the total propensity for the reactions to occur,

$$a_0 = \sum_{\mu=1}^{M} a_\mu\,.$$

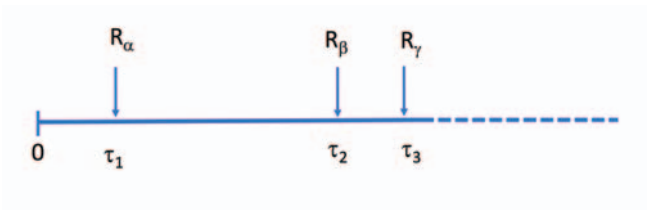

Figure A. Given the configuration at time t=0, one computes a_0 and generates an exponentially distributed random number to get the time τ_1, namely the time for the first reaction. The first reaction is determined to be R_α (see *Figure* B). Carrying this out will change the configuration, and with this changed configuration, one recomputes a_0 to give the next time, τ_2 when reaction R_β occurs, and so on.

Continued

Box 1. *Continued*

The probability that reaction ν occurs after time τ is therefore:

$$P(\tau, \nu) = a_\nu P_0(\tau).$$

In other words, no reaction occurs for time τ, and then reaction ν occurs. It is clear, therefore, that the random variable τ, namely the time *between* reactions follows an exponential distribution with rate a_0 and this leads to the following straightforward algorithm

- Step 1: Given a configuration, compute the propensity for each chemical reaction, namely the factors $a_\mu, \mu = 1, \ldots, M$ and therefore also a_0.
- Step 2: Generate the uniform random number r_1 in the interval [0,1].
- Step 3: The next reaction will take place at $\tau = (-\ln r_1)/a_0$ (τ thus has the required exponential distribution). A cartoon of this procedure is shown in *Figure* A (see caption).
- Step 4: Generate another uniform random number r_2, also in the interval [0,1].
- Step 5: Use r_2 to determine which reaction will take place as follows. Find ν such that

$$\sum_{\mu=1}^{\nu-1} a_\mu < r_2 a_0 \leq \sum_{\mu=1}^{\nu} a_\mu.$$

  (See *Figure* B for an illustration of how this is done.) Then reaction R_ν will take place after time τ, and this means that the configuration should be changed accordingly. Using the stoichiometry of the reaction R_ν, it is required that the numbers of the reactant species should be decreased, and correspondingly, the numbers of the product species be increased.

- Step 6: Return to Step 1 with the changed configuration, having advanced the time by τ.

Continued

Box 1. *Continued*

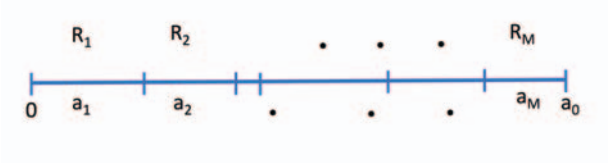

Figure B. Since one of the M reactions must occur, and the total propensity is a_0, one simply lines up the different reactions $R_1, R_2, \ldots, R_M$ and randomly selects one of these with probability proportional to its relative propensity.

Suggested Reading

[1] **E L King, *How Chemical Reactions Occur. An Introduction to Chemical Kinetics and Reaction Mechanisms*, Benjamin, New York, 1963.**

[2] **K J Laidler, *Chemical Kinetics, (3/e)*, Pearson, 2003.**

[3] **H A Kramers, Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions, *Physica*, Vol.7, 284, 1940.**

[4] **M Delbrück, Statistical Fluctuations in Autocatalytic Kinetics, *J. Chem. Phys*., Vol.8, pp.120–124, 1940.**

[5] **D A McQuarrie, Stochastic Approach to Chemical Kinetics, *J. Appl. Probab*., Vol.4, p.413, 1967.**

[6] **D T Gillespie, Exact Stochastic Simulation of Coupled Chemical Reactions, *J. Phys. Chem*., Vol.81, pp.2340–2361, 1977.**

[7] **Y Cao and D C Samuels, Discrete Stochastic Simulation Methods for Chemically Reacting Systems, *Methods Enzymol*., Vol.454, pp.115–140, 2009; T Székely and K Burrage, Stochastic Simulation in Systems Biology, *Comp. and Struct. Biotech. J*., Vol.12, pp.14–25, 2014.**

[8] **F Schögl, Chemical Reaction Models for Non-equilibrium Phase Transition, *Z. Physik*., Vol.253, pp.147–161, 1972.**

[9] **H McAdams and A Arkin, Stochastic Mechanisms in Gene Expression, *Proc. Natl. Acad. Sci*., USA, Vol.94, pp.814–819, 1997.**

[10] **B C Goodwin, Oscillatory Behaviour in Enzymatic Control Processes, *Adv. Enzyme Regul*., Vol.3, pp.425–438, 1965.**

[11] **J S Griffith, Mathematics of Cellular Control Processes: Negative Feedback to One Gene, *J. Theor. Biol*., Vol.20, 202, 1968.**

[12] **J M Vilar, H Y Kueh, N Barkai and S Leibler, Mechanisms of Noise-resistance in Genetic Oscillators, *Proc. Natl. Acad. Sci*., USA, Vol.99, pp.5988–5992, 2002.**

[13] **G Balázsi, A van Oudenaarden and J J Collins, Cellular Decision-making and Biological Noise: From Microbes to Mammals, *Cell*, Vol.144, pp.910–925, 2011.**

[14] D Bernstein, Simulating Mesoscopic Reaction-diffusion Systems Using the Gillespie Algorithm, *Phys. Rev. E*., Vol.71, p.041103, 2005; X Cai, Exact Stochastic Simulation of Coupled Chemical Reactions with Delays, *J. Chem. Phys*., Vol.126, p.124108, 2007.

Address for Correspondence
Ashwin B R Kumar
Department of Biochemistry
and Biophysics
University of Rochester
Medical Center
Rochester, NY 14642, USA.
Email: ashwin_kumar@urmc.
rochester.edu
Ram Ramaswamy
School of Physical Sciences
Jawaharlal Nehru University
New Delhi 110 067, India.
Email: ramaswamy@jnu.ac.in