

# DETERMINISTIC AND STOCHASTIC SIMULATION AND ANALYSIS OF BIOCHEMICAL REACTION NETWORKS: THE LACTOSE OPERON EXAMPLE

Necmettin Yildirim\* and Caner Kazanci†

## Contents

1. Introduction	372
2. Mathematical Modeling of Biochemical Reaction Networks and Law of Mass Action	372
2.1. Simple enzymatic reactions and Michaelis–Menten equation	374
2.2. Higher order kinetics and Hill equations	375
2.3. Steady state and linear stability analysis in one-dimensional models	377
2.4. Modeling coupled reactions and bistability	378
3. Stochastic Simulations	381
3.1. Cases where stochasticity matters	382
3.2. Stochastic simulation algorithms	384
4. An Example: Lactose Operon in <i>E. coli</i>	386
5. Conclusions and Discussion	393
Acknowledgment	395
References	395

## Abstract

A brief introduction to mathematical modeling of biochemical regulatory reaction networks is presented. Both deterministic and stochastic modeling techniques are covered with examples from enzyme kinetics, coupled reaction networks with oscillatory dynamics and bistability. The Yildirim–Mackey model for lactose operon is used as an example to discuss and show how deterministic and stochastic methods can be used to investigate various aspects of this bacterial circuit.

\* Division of Natural Sciences, New College of Florida, Sarasota, Florida, USA

† Department of Mathematics/Faculty of Engineering, University of Georgia, Athens, Georgia, USA

## 1. INTRODUCTION

Recent advancements in experimental techniques in biology and medicine enable high-throughput experiments acquire data that are easier, cheaper, and more accurate. As more data became available, the need for the right mathematical and computational tools for analysis and interpretation became clear. Until recently, reductionist methods and statistics were the main tools used to study biological systems. “Simpler diseases” were understood and treated this way, but the remaining complicated ones, such as cancer, AIDS, sepsis, and heart diseases require new sophisticated techniques that can cope with their complexity. Structure of biological organisms enables researchers to use various mathematical and computational approaches for modeling, simulation, and analysis. At cellular level, there are distinct biochemical mechanisms that are responsible for specific jobs, such as energy production, protein synthesis, motor functions, defense, and signaling.

We present a brief introduction to deterministic and stochastic modeling of biochemical networks. Outline of this chapter is as follows: In Section 2, we describe the basics of mathematical modeling of biochemical reactions, their steady state and stability analysis using the law of mass action. Section 3 summarizes stochastic modeling techniques and describes a basic stochastic algorithm. We focus on the Yildirim–Mackey model for *lac* operon in Section 4 and discuss how bistability arises in this network. We then simulate the *lac* operon model using both the deterministic and stochastic methods with experimentally estimated parameter values to show that this system indeed displays bistable behavior for physiologically reasonable parameters values. The chapter ends with Section 5 which includes conclusions and discussion.

## 2. MATHEMATICAL MODELING OF BIOCHEMICAL REACTION NETWORKS AND LAW OF MASS ACTION

There are different approaches and methodologies to studying biochemical reactions. Mass-action kinetics which results in system of differential equations are commonly used to describe the dynamics of biochemical reaction networks. This approach is a fully deterministic and it is appropriate when a system under consideration has large number of molecules and these molecules are spatially homogeneous. In this section, we briefly describe how to construct differential equation models that describe dynamics of a reaction network under these two assumptions.

Suppose that  $A$ ,  $B$ , and  $C$  are three proteins and when molecules of  $A$  collide with molecules of  $B$ , they may react and form  $C$ . Assume that this reaction is associated with a positive rate constant  $k_1$ , quantifying how likely

it is for such a collusion to result in a reaction. We also assume that  $C$  can break into  $A$  and  $B$ . Let us assume that the backward positive rate constant for reaction is  $k_2$ . We use the following notation Eq. (12.1) to denote this chemical reaction system:



Now suppose that there exist various amounts of these proteins in a constantly well-stirred pot, so that its contents remain spatially homogeneous. Here, we are interested in the temporal evolution of the molecular concentrations of these protein molecules. Let us denote the concentration of a species  $X$  with  $[X]$ . We would like to construct a system of differential equations that governs the temporal evolution of  $[A]$ ,  $[B]$ , and  $[C]$ . Naturally, we can think of the reaction given in Eq. (12.1) as two separate reactions:  $A + B \xrightarrow{k_1} C$  and  $C \xrightarrow{k_2} A + B$ .

According to mass-action kinetics, the time derivative of the concentration of protein  $A$  is equated to the difference between the sum of the gain terms (input chemical fluxes) that cause the concentration to increase and the sum of the loss terms (output chemical fluxes) that act to decrease the concentration as:

$$\frac{d[A]}{dt} = \sum \text{Input fluxes} - \sum \text{Output fluxes}. \quad (12.2)$$

For the reaction given in Eq. (12.1), the mathematical model is

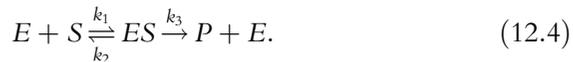
$$\begin{aligned} \frac{d}{dt}[A] &= \frac{d}{dt}[B] = -k_1[A][B] + k_2[C], \\ \frac{d}{dt}[C] &= k_1[A][B] - k_2[C]. \end{aligned} \quad (12.3)$$

This system of differential equations can be solved to simulate temporal evolution of  $[A]$ ,  $[B]$ , and  $[C]$  after assigning values for initial concentrations of  $A$ ,  $B$ , and  $C$  for  $t = 0$ . Although mass-action kinetics is extremely useful for modeling chemical reactions, biological systems benefit greatly from enzymatic kinetics. Most chemical reactions in biological organisms rely on enzymes, special molecules that enable certain reactions to occur. In general, enzymes are fast acting molecules existing in low concentrations. These properties enables derivation of simpler equations for enzymatic reactions using certain approximations. In the following two sections, we give two examples from enzyme kinetics. In the first example, the rate for the product formation is a hyperbolic function of the substrate

concentration. In the second example, the rate of product is a sigmoidal function of the substrate concentration. Then we discuss importance of these types of functional relationships.

## 2.1. Simple enzymatic reactions and Michaelis–Menten equation

Consider the following enzyme catalyzed reaction given in Eq. (12.4). An enzyme  $E$  binds to a substrate  $S$  and forms an enzyme–substrate complex  $ES$  with a rate constant  $k_1$ . We assume this reaction is fully reversible. That is to say  $ES$  can break down into  $E$  and  $S$ , and suppose that the associated rate constant for this backward reaction is  $k_2$ . In this reaction network, we assume that  $E$  can also release from  $ES$  and produce  $P$  and  $E$ . The rate constant for this final step of the reaction is  $k_3$ . In this simple system, there are four time dependent variables:  $[E]$ ,  $[S]$ ,  $[ES]$ , and  $[P]$ .



We assume that the total concentrations of the enzyme and the substrate stay constant over time for this system. That gives us the following two equations:

$$E_0 = [E] + [ES], \quad (12.5)$$

$$S_0 = [S] + [ES] + [P], \quad (12.6)$$

where  $E_0$  and  $S_0$  are the initial concentrations of the enzyme and the substrate, respectively. These two equations reduce the number of free variables from four to two. Now, we can write two differential equations that describe the dynamics of the concentration of  $ES$  and  $P$ .

$$\frac{d[ES]}{dt} = k_1[E][S] - (k_2 + k_3)[ES]. \quad (12.7)$$

$$\frac{d[P]}{dt} = k_3[ES]. \quad (12.8)$$

These equations describe the dynamics of the single enzyme–substrate reaction in Eq. (12.4). However, we can further simplify this model using additional assumptions. Not all variables in a dynamic system change at same time scale. It is often the case that some variables change significantly faster than others. If we assume  $[ES]$  is a fast variable and reaches a steady state much earlier than  $[P]$ , then we get  $d[ES]/dt \approx 0$  and hence

$$[ES] = \frac{k_1}{k_2 + k_3} [E][S]. \quad (12.9)$$

This is called the *quasi steady state assumption* on  $[ES]$ . If we substitute  $[ES]$  given in Eq. (12.9) into Eq. (12.5) and solve the resultant equation for  $[E]$ , we get

$$[E] = \frac{E_0}{1 + (k_1/k_2 + k_3)[S]}. \quad (12.10)$$

Plugging Eq. (12.9) into Eq. (12.8) after replacing  $[E]$  in Eq. (12.9) by Eq. (12.10) gives us

$$\frac{d[P]}{dt} = \frac{V_{\max}[S]}{K_m + [S]}, \quad (12.11)$$

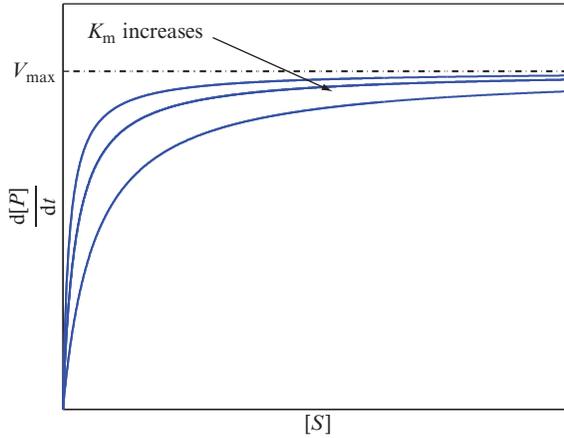
where  $V_{\max} = k_3 E_0 = k_2 + k_3 / k_1$ . This equation is well known as Michaelis–Menten equation in enzyme kinetics.

In Eq. (12.11), the parameter  $V_{\max}$  is the maximum rate that this reaction can occur and the parameter  $K_m$  is defined as the values of  $[S]$  that gives half of  $V_{\max}$ . In other words, when  $[S] = K_m$ , the product formation ( $d[P]/dt \approx -d[S]/dt$ ) occurs at half of its maximum rate ( $V_{\max}$ ). A graphical representation of Michaelis–Menten equation in Eq. (12.11) for various values of  $K_m$  is depicted in Fig. 12.1. As seen in this graph, all the curves approach to a maximum value of  $V_{\max}$  as  $[S]$  increases. For larger  $K_m$  values, the curves shift toward the right. All the curves are concave downward and the concavities of these curves do not change as  $[S]$  increases.

## 2.2. Higher order kinetics and Hill equations

Consider the following reaction system given in Eq. (12.12). In this reaction system,  $n$ -molecules of a substrate  $S$  bind to an enzyme  $E$  and form a complex  $ES_n$  with a forward rate constant  $k_1$  and a reverse rate constant  $k_2$ . Then the enzyme is released and a product  $P$  is formed with a rate constant  $k_3$





**Figure 12.1** A graphical representation of Michaelis–Menten equation in Eq. (12.11) and hyperbolic kinetics for various values of  $K_m$  when the maximum rate  $V_{\max}$  kept fixed. As  $K_m$  increases, the curves move to the right and all curves are looking downward.

Let us also assume that at any time throughout the course of these reactions, the first reaction (Eq. (12.12)) is much faster compared to the second one (Eq. (12.13)). Therefore the first reaction reaches equilibrium (forward and backward reaction rates become equal) earlier than  $P$  starts to get produced. This allows us to write

$$\frac{d[ES_n]}{dt} = k_1[E][S]^n - k_2[ES_n] = 0.$$

This is called the *equilibrium assumption*. Then, we can write

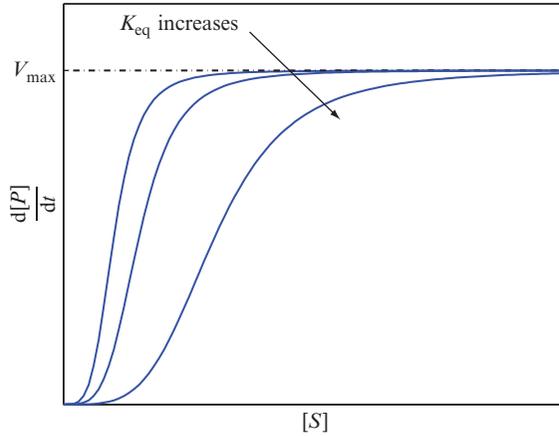
$$[E] = \frac{k_2[ES_n]}{k_1[S]^n}. \quad (12.14)$$

Assuming that the total amount of enzyme is conserved, we can write

$$E_{\text{tot}} = [E] + [ES_n]. \quad (12.15)$$

Substituting Eq. (12.14) into Eq. (12.15) and then solving it for  $[ES_n]$ , we obtain

$$[ES_n] = \frac{E_{\text{tot}}[S]^n}{K_{\text{eq}} + [S]^n}, \quad K_{\text{eq}} = \frac{k_2}{k_1}.$$



**Figure 12.2** A graphical representation of Hill Equation in Eq. (12.16) and sigmoidal kinetics for various values of  $K_{eq}$  when the maximum rate  $V_{max}$  kept fixed for  $n = 4$ . Similar to the Michaelis–Menten curves, as  $K_{eq}$  values increase, the curves move to the right. Unlike Michaelis–Menten curves, the Hill function curves are concave upward for smaller values of  $[S]$  and then they become concave downward for larger values of  $[S]$ .

According to mass-action law, the dynamics of the concentration of  $P$  is proportional to the concentration of  $[ES_n]$  with proportionality constant  $k_3$ . Hence,  $d[P] / dt$  takes the following form

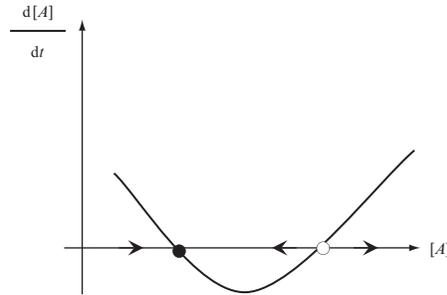
$$\frac{d[P]}{dt} = \frac{V_{max}[S]^n}{K_{eq} + [S]^n}, \tag{12.16}$$

where,  $V_{max} = k_3 E_{tot}$  and  $K_{eq} = k_2 / k_1$ . A graphical representation of Hill equation in Eq. (12.16) for various values of  $K_{eq}$  when  $n = 4$  is shown in Fig. 12.2. As seen in this figure, all the curves approach to a maximum value of  $V_{max}$  as  $[S]$  increases. For larger  $K_{eq}$  values, the curves shift toward the right. One of the important characteristics of Hill equation curves is that they are all concave upward and then become concave downward after a threshold value of  $[S]$  for no matter what  $K_{eq}$  values are. This important feature leads to bistability, ability of a system to rest in two steady states, as will be discussed in the following sections.

### 2.3. Steady state and linear stability analysis in one-dimensional models

One-dimensional mathematical model has the following general form

$$\frac{d[A]}{dt} = f([A]). \tag{12.17}$$



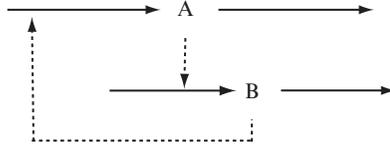
**Figure 12.3** Graphical approach to the steady state and stability analysis of one-dimensional model given in Eq. (12.17).

We say a point  $[A^*]$  is a *steady state* if time derivative at that point is zero. The steady states can be computed by solving the equation  $f([A^*]) = 0$ , since  $d[A]/dt = 0$  at  $[A] = [A^*]$ . When Eq. (12.17) is plotted as a function taking  $[A]$  on  $x$ -axis and  $d[A]/dt$  on  $y$ -axis, the  $x$ -intercepts gives the steady state values (see Fig. 12.3). Now, we can think of the equation given by Eq. (12.17) a model that describes movement of an imaginary particle moving along  $[A]$ -axis and  $d[A]/dt$  is the velocity of that particle. Since  $d[A]/dt = 0$  at the steady state value, there is no change in  $[A]$  when  $d[A]/dt = 0$ . If  $d[A]/dt < 0$  for a value of  $[A]$ , the arrows point to the left, otherwise they point to the right. As can be seen in Fig. 12.3, there are two types of steady states. The filled dot represents the *stable* steady state since the flow is toward this steady state. The open circle represent the *unstable* steady state since the flow is toward away from this steady state.

We can conclude from Fig. 12.3 that a steady state is stable if  $d[A]/dt < 0$  at that steady state value. It is an unstable steady state if  $d[A]/dt > 0$  holds at that steady state value.

## 2.4. Modeling coupled reactions and bistability

In this section, we give an example with a positive feedback loop, one of the important regulatory mechanisms in biological systems. It is capable of producing two stable steady states separated by an unstable steady state, so called “bistable system.” Bistability provides a true discontinuous switching between stable steady states. A bistable system often involves a positive feedback loop. Positive feedback loops are ubiquitous control mechanisms in gene networks. The lactose operon and the arabinose operon of *Escherichia coli* are two examples of this type of regulatory control networks (Lewin, 2008; Schleif, 2000). Consider the hypothetical system with positive feedback loop in Fig. 12.4. This reaction network has two proteins  $A$  and  $B$ . We use the Eqs. (12.18) and (12.19) to model the dynamics of this



**Figure 12.4** A cartoon for a reaction network with a positive feedback loop.

toy reaction network with positive feedback. The dynamics of the concentration of  $A$  is given by Eq. (12.18). In this network, we assume  $A$  is produced with a constant rate  $\alpha_1$  and degraded at a rate proportional to its concentration with a proportionality constant  $\beta_1$ . The second term in this equation is for the increase in the production rate of  $A$  due to the positive feedback and we assume that this relationship has a Hill function form with  $n = 2$ . Equation (12.19) models the dynamics of  $[B]$ . It is assumed that  $A$  is required for the production of  $B$  and the production rate of  $A$  is proportional to the concentration of  $A$  with a proportionality  $\alpha_2$ . We assume that  $B$  has a decay constant  $\beta_2$ .

$$\frac{d[A]}{dt} = \alpha_1 + \frac{V_m[B]^2}{K_m + [B]^2} - \beta_1[A]. \quad (12.18)$$

$$\frac{d[B]}{dt} = \alpha_2[A] - \beta_2[B]. \quad (12.19)$$

This system of differential equations has two time dependent variables  $[A]$  and  $[B]$  and six positive parameters  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$ ,  $\beta_2$ ,  $V_m$ , and  $K_m$ .

#### 2.4.1. Steady state and stability analysis

Suppose that the system given in Eqs. (12.18) and (12.19) has a steady state  $([A^*], [B^*])$ . At this steady state  $([A^*], [B^*])$ ,  $d[A]/dt = d[B]/dt = 0$  has to hold simultaneously. Therefore, we can write

$$\alpha_1 + \frac{V_m[B^*]^2}{K_m + [B^*]^2} - \beta_1[A^*] = 0, \quad (12.20)$$

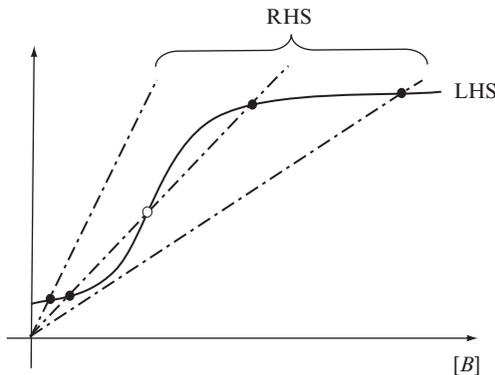
$$\alpha_2[A^*] - \beta_2[B^*] = 0. \quad (12.21)$$

After solving Eq. (12.21) for  $[A^*]$  and plugging it back in Eq. (12.20), we get a nonlinear equation in  $[B^*]$  as

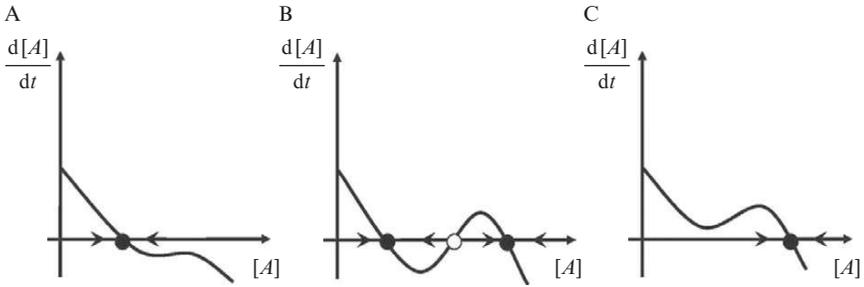
$$\alpha_1 + \frac{V_m[B^*]^2}{K_m + [B^*]^2} = \beta_1 \frac{\beta_2}{\alpha_2} [B^*]. \quad (12.22)$$

When we look at the right-hand side and left-hand side of Eq. (12.22) as two different functions of  $[B]$ , we see that the right-hand side is a linear function of  $[B]$  with a positive slope  $m = \beta_1(\beta_2 / \alpha_2)$ . The left-hand side is a Hill function which equals  $\alpha_1$  when  $[B] = 0$ , and approaches to a maximum value of  $\alpha_1 + V_m$  as  $[B] \rightarrow \infty$  (Fig. 12.5). When both of these functions are plotted in the same plane choosing  $x$ -axis as  $[B]$ , where these two functions intersect each other give us the steady state value or values. It is not hard to see that these two functions can intersect each other at only one point for small and large values of  $m$  and but at three points for intermediate values of  $m$ . Since the left-hand side of Eq. (12.22) is a Hill function and it is concave upward for small values of  $[B]$  that means increase in this function is relatively small for smaller concentration of  $B$ . Then there is a sharp increase for intermediate concentration of  $B$  and after that the curve changes its concavity and becomes concave downward and finally it levels off. This feature of the curve allows possibility of having multiple steady states for intermediate values of  $m$ . Figure 12.5 shows how one, two, or three steady states can arise in this model for different values of  $m$ .

The local stability analysis of the model given in Eqs. (12.18) and (12.19) can be studied mathematically by linearizing this system of differential equations around a given steady state and looking at eigenvalues of the jacobian matrix. For the sake of simplicity, let us assume  $[B]$  is a fast variable in this system and  $d[B] / dt = 0$  in Eq. (12.19). After solving  $d[B] / dt = 0$  in Eq. (12.19) for  $[B]$  and putting it into Eq. (12.18), the two-dimensional model reduces to a one-dimensional model as:



**Figure 12.5** A diagrammatic representation showing how one, two, or three steady states may arise in the model given by Eqs. (12.18) and (12.19). The solid line represents the left-hand side of Eq. (12.22). The dash-dotted lines are for the right-hand side of Eq. (12.22) for three different values of  $m = \beta_1\beta_2 / \alpha_2$ . As seen in this plot, there is only one steady state for either small or large values of  $m$ . However, there is a range for  $m$  in which it is possible to have three coexisting steady states.



**Figure 12.6** A diagrammatic representation of stability analysis of the bistable system modeled by Eq. (12.23). In this plot we see that when there is only one steady state (A and C), this steady state is stable (the flow is toward this steady state). When there are three steady states (B), the middle one is unstable (the flow is away from this steady state) and the other two are stable.

$$\frac{d[A]}{dt} = \alpha_1 + \frac{V_m[A]^2}{K_m(\beta_2/\alpha_2)^2 + [A]^2} - \beta_1[A]. \quad (12.23)$$

In Fig. 12.6, we plot  $d[A]/dt$  versus  $[A]$  for small, medium, and large values of  $m$ . As seen in this figure, there is only one steady state for small and large values of  $m$  and these steady states are always stable (A and C in Fig. 12.6). When there exists three steady states (B in Fig. 12.6), the middle steady state is always unstable and the lowest and the highest steady states are always stable.

### 3. STOCHASTIC SIMULATIONS

Ordinary differential equation (ODE) models are widely used to simulate biochemical reaction systems. However, they are by no means a perfect in capturing every aspect of molecular reactions that occur in real life. Continuous variables used in ODEs are not appropriate to represent dynamics of molecular species that exist in low quantities in a system. Another major shortcoming of ODEs shows up in systems capable of multiple steady states. The deterministic solution of the ODE representation of such a system will always converge to a single stable steady state and stay there. However, in real life, constant switching behavior among steady states may be observed. Due to inherent fluctuations within the system, the state may be “pushed” from one steady state to another. Such issues may be accommodated by stochastic simulations, or stochastic differential equations (SDE).

### 3.1. Cases where stochasticity matters

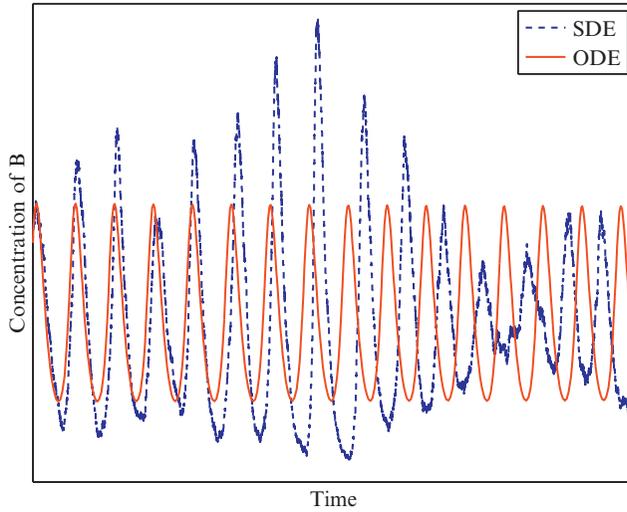
Molecules in a medium move and collide with each other. When two molecules of the same species collide, their speed and velocity changes. However, when two molecules with the capability to react collide, they may react with some probability  $p$  and form a new chemical species. This probability  $p$  is somewhat analogous to the reaction rate  $k$  in mass-action kinetics. This inherent probabilistic behavior may not be captured by deterministic differential equations. In general, the effects of this probabilistic behavior may be ignored without much penalty. However, in some cases, significant difference is observed between the stochastic and deterministic equations representing the same system. Here, we give two such example systems. For biochemical reaction systems, an ODE solution is an approximation to a stochastic phenomenon. For example, at equilibrium, the time course plot for the ODE solution is a straight horizontal line, while a continuous “noisy” activity is observed for the SDE solution. In reality, unlike the ODE simulation suggests, the activity in a biochemical reaction system never stops. At equilibrium, molecules keep colliding and the reactions keep occurring, but at a balanced rate so that the molecular species concentrations stay the same on average.

The difference between the stochastic and deterministic simulations become significant as the number of molecules decrease, in which case, the noise takes over the dynamics. We demonstrate this effect of probabilistic behavior on the following reaction network (Scott, 1991):



In this system, a constant supply of  $A$  is assumed.  $B$  is a product, so its concentration does not affect the system. The only changing quantities in the two-dimensional ODE are the concentrations of  $X$  and  $Y$ . Initial conditions are  $[X](0) = [Y](0) = 300$ , and the reaction rates are  $k_1[A] = 2$ ,  $k_2 = 0.01$ ,  $k_3 = 2$ . This reaction system is chosen because the concentrations of  $X$  and  $Y$  oscillates, clearly demonstrating how the inherent probabilistic behavior of molecular reactions may perturb the dynamics predicted by the ODE simulation. In Fig. 12.7, we compare the ODE and SDE simulations for this reaction system. Significant variation is observed between the two methods. This difference is expected to be more apparent in case of smaller systems, where there are small number of molecules present in the medium. In extreme cases, the dynamic behavior may be totally lost and dominated by noise.

Even if the number of molecules remains high in the environment at all times, the ODE simulation still may convey a significantly different



**Figure 12.7** The difference between the ODE and SDE simulations for the same reaction system (Eq. (12.24)) is demonstrated. The time course of the concentration of  $B$  is shown to compare the two methods. The ODE solution for  $[B]$  converges to a limit cycle over time and stays there. Although the SDE solution initially shows similar behavior, it quickly diverges and displays significant variation in both amplitude and phase.

behavior than a stochastic simulation. The positive feedback mechanism given in Fig. 12.4 demonstrates such behavior.

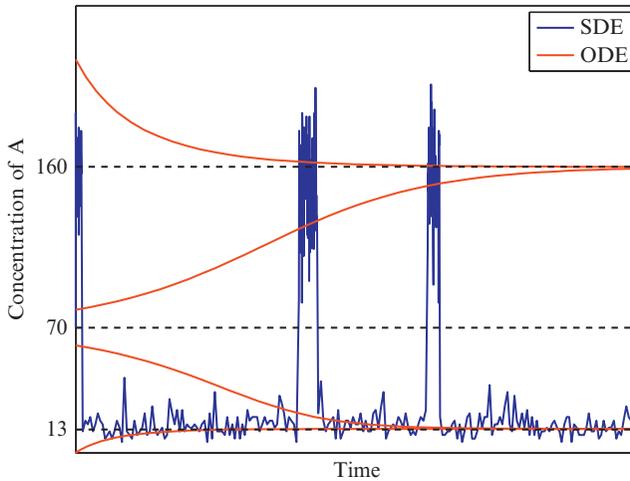
Although this system consists of two molecular species, it is possible to derive a one-dimensional ODE using the quasi steady state assumption for  $B$  as in Eq. (12.23) under the assumption given by Eq. (12.21). Note that this function is capable of having multiple steady states. For the following choice of parameters:

$$\alpha_1 = 0.42, \quad \beta_1 = 0.004, \quad \alpha_2 = 0.36, \quad \beta_2 = 32, \quad V_m = 0.93, \quad K_m = 1.8,$$

the system has two stable steady states at 13, 160 and an unstable steady state at 70. Depending on the initial condition, the ODE solution converges to one of the stable steady states (Fig. 12.8):

$$\lim_{t \rightarrow \infty} [A](t) = \begin{cases} 13, & \text{if } 0 \leq [A](0) < 70, \\ 70, & \text{if } [A](0) = 70, \\ 160, & \text{if } 70 < [A](0). \end{cases}$$

Stochastic simulation of this bistable system conveys an interesting behavior; it switches back and forth between the two stable steady states. The switching occurs when the inherent perturbations around a steady state are large enough to push the solution to the other side of the unstable steady



**Figure 12.8** A stochastic solution of the positive feedback system in Eq. (12.23) with initial condition at  $[A](0) = 70$  is plotted, which conveys a “switching” behavior between the stable steady states. Steady states are shown with dashed lines.

state. This phenomenon continues repeatedly, preventing the system to stay at one steady state. This is the case even if  $A$  is present in abundant concentrations. Therefore what happens here is a lot more than a stochastic solution showing noisy behavior. If a system has multiple steady states, the stochastic solution may show significantly different results than the ODE solution. Stochastic methods may be a necessity for such systems.

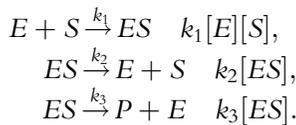
### 3.2. Stochastic simulation algorithms

Like numerical ODE solutions, various methods exist for stochastic simulations. Unlike numerical ODE solutions, there are many other factors to consider in choosing the correct stochastic method. Note that it is not possible to add some Gaussian noise to an ODE solution at each iteration to get a stochastic solution. Result of such an approach would be noisy and wrong. Similar but correct approaches exist, such as chemical Langevin equation (Gillespie, 2000), where the correct noise term is computed and added to the ODE solution. This method was used in Fig. 12.7. Computation of the correct noise term is essential, and is more complicated than computing the deterministic part of the solution. Chemical Langevin equation is a first order method. Higher order methods exist. However, computation of the noise term get extremely complicated. Another issue with this approach occurs when some molecular species exist in extremely low concentrations, in which case, the solution may go negative, indicating negative

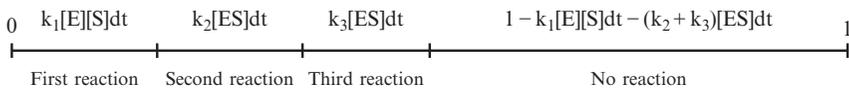
concentrations. This implausible result can occur because the noise term may exceed an extremely small deterministic component of the solution.

A different well-known methodology is proposed by Gillespie (1977), where each individual molecule and reaction is taken into account. Therefore it works very well even if some molecular species exist in low concentrations. However, if some molecular species exist in extremely high concentrations, then Gillespie’s stochastic algorithm may run extremely slow. There are many recent developments aimed to eliminate the mentioned shortcomings of stochastic solutions. Various modifications of Gillespie’s stochastic method that run faster (Gillespie, 2001; Gillespie and Petzold, 2003), and chemical Langevin equations that preserve positivity (Wilkie and Wong, 2008) are being developed, though such enhancements generally come at the cost of another compromise such as accuracy, complexity, or efficiency.

We will go over a basic stochastic simulation method for demonstration purposes. Although Gillespie’s algorithm or chemical Langevin equations are considered better performing methods in general, basic stochastic algorithm is simple and intuitive. We will use the reaction system given in Eq. (12.4) to describe the basic stochastic algorithm. The three reactions and their associated reaction rates are as follows:



The method is based on the probabilities that these reactions occur over a fixed time interval  $dt$ . The smaller this time interval, the more accurate the simulation. We compute the probability that one of these reactions occurs during a time interval of length  $dt$  by multiplying its reaction rate with  $dt$ . For example, the probability that the first reaction will occur during a time interval of length  $dt$  is  $k_1[E][S]dt$ . Here,  $dt$  has to be sufficiently small, so that this product stays less than 1. Actually,  $dt$  is supposed to be small enough that at most one of these reactions happen during  $dt$ . The probability that none of these reactions will occur over  $dt$  is  $1 - k_1[E][S]dt - (k_2 + k_3)[ES]dt$ . Then we can devise an iterative simulation algorithm where we update the state of the system after a fixed time interval of length  $dt$ . We decide which reaction occurs by partitioning the interval  $[0, 1]$  into four subintervals with lengths equal to the corresponding probabilities:



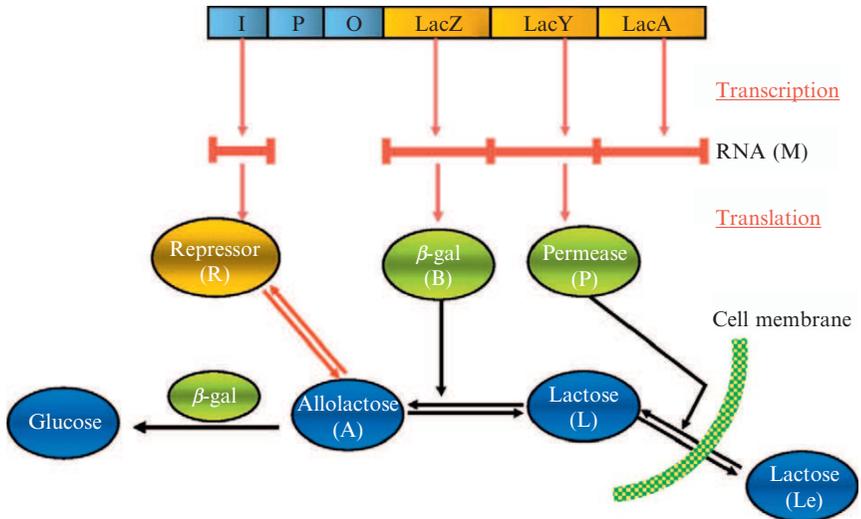
We then choose a random value  $p$  between 0 and 1, and we update the state of the system depending on which subinterval  $p$  belongs to. For example, if  $0 < p < k_1[E][S]dt$ , then it means that  $E + S \rightarrow ES$  has occurred once, therefore we update the system so that  $E$  and  $S$  lose one molecule each and  $ES$  gains one. Or if  $k_1[E][S]dt + (k_2 + k_3)[ES]dt < p < 1$  then it means that no reaction has occurred, so we jump to the next iteration without changing the state. This iterative scheme can be generalized for any chemical reaction system. At each iteration:

1. Reaction probabilities are computed.
2. Interval  $[0,1]$  is partitioned into subintervals according to these probabilities.
3. A uniform random number  $p$  is chosen from  $[0,1]$ .
4. System state is updated depending on which subinterval  $p$  belongs to.

Similar to numerical ODE solvers, the accuracy of this algorithm increases as  $dt$  decreases. Similarly, if  $dt$  is extremely small, then simulations will take longer computing time. For smaller  $dt$  values, system state will not change during many iterations because the probability that no reaction occurs converges to 1 as  $dt \rightarrow 0$ . Gillespie's stochastic algorithm provides a remedy for this issue by choosing an adaptive time-step  $dt$ , which is explained in his paper (Gillespie, 1977).

#### 4. AN EXAMPLE: LACTOSE OPERON IN *E. COLI*

We use the lactose operon (the *lac* operon) of *E. coli* and a modified version of the Yildirim–Mackey model (Mackey *et al.*, 2004; Yildirim and Mackey, 2003; Yildirim *et al.*, 2004) developed for this bacterial regulatory circuit to demonstrate the methods and analysis described in previous sections. The *lac* operon is the classical example of an inducible circuit which encodes the genes for the transport of external lactose into the cell and its conversion to glucose and galactose. A cartoon that depicts the major components of this circuit is shown in Fig. 12.9. The molecular mechanism of the *lac* operon works as follows: The *lac* operon has a small promoter/operator region ( $P$  and  $O$ ) and three larger structural genes *lacZ*, *lacY*, and *lacA*. There is a regulatory gene *lacI* preceding the *lac* operon. *lacI* is responsible for producing a repressor ( $R$ ) protein. In the presence of allolactose, a binary complex is formed between allolactose and the repressor that makes binding of the repressor to the operator region impossible. In that case, the RNA polymerase bound to the promoter is able to initiate transcription of the structural genes to produce mRNA( $M$ ). However, in the absence of allolactose ( $A$ ) the repressor protein  $R$  binds to the operator region  $O$  and prevents the RNA polymerase from transcribing the structural



**Figure 12.9** Schematic representation of the lactose operon regulatory system. See the text for details.

genes. Once the mRNA has been produced, the process of translation starts. The *lacZ* gene encodes the portion of the mRNA that is responsible for the production of  $\beta$ -galactosidase (*B*) and translation of the *lacY* gene produces the section of mRNA that is ultimately responsible for the production of an enzyme permease (*P*). The final portion of mRNA produced by transcription of the *lacA* gene encodes for the production of thiogalactoside transacetylase which is thought not to play a role in the regulation of the *lac* operon (Beckwith, 1987). This positive control system works as follows: When there is no glucose available for cellular metabolism but if lactose (*L*) is available in a media, the lactose is transported into the cell by the permease. This intracellular lactose is then broken down into glucose, galactose, and allolactose by  $\beta$ -galactosidase. The allolactose is also converted to glucose and galactose by the same enzyme  $\beta$ -galactosidase. The allolactose feeds back to bind with the lactose repressor and enable the transcription process which completes the positive feedback loop.

Yildirim *et al.* (Mackey *et al.*, 2004; Yildirim and Mackey, 2003) devised a mathematical model which takes into account the dynamics of the permease, internal lactose,  $\beta$ -galactosidase, the allolactose interactions with the *lac* repressor, and mRNA. The final model consists of five nonlinear differential delay equations with delays due to the transcription and translation process. We modified this model in this study and eliminated the delay terms. This change reduced the original model to a five-dimensional system of ODEs. The equation of this model are given in Eqs. (12.25)–

(12.29). The estimated values for the model parameters from the published data are listed in Table 12.1. The details on the development of this model and estimation of the parameters can be found in Mackey *et al.* (2004), Yildirim and Mackey (2003), Yildirim *et al.* (2004) (Table 12.2).

We studied this model using both deterministic and stochastic approaches. To see if the modified model captures the experimentally

**Table 12.1** The model parameters estimated from experimental data (from Yildirim and Mackey, 2003)

$n$	2	$\mu_{\max}$	$3.47 \times 10^{-2} \text{ min}^{-1}$
$\gamma_M$	$0.411 \text{ min}^{-1}$	$\gamma_B$	$8.33 \times 10^{-4} \text{ min}^{-1}$
$\gamma_A$	$0.52 \text{ min}^{-1}$	$\Gamma_0$	$7.25 \times 10^{-7} \text{ mM/min}$
$K$	7200	$\alpha_M$	$9.97 \times 10^{-4} \text{ mM/min}$
$K_{L_1}$	1.81 mM	$\alpha_A$	$1.76 \times 10^4 \text{ min}^{-1}$
$K_A$	1.95 mM	$\alpha_B$	$1.66 \times 10^{-2} \text{ min}^{-1}$
$\gamma_L$	$0.0 \text{ min}^{-1}$	$\beta_A$	$2.15 \times 10^4 \text{ min}^{-1}$
$\alpha_L$	$2880 \text{ min}^{-1}$	$K_L$	$9.7 \times 10^{-4} \text{ M}$
$K_{L_e}$	0.26 mM	$\gamma_P$	$0.65 \text{ min}^{-1}$
$\beta_{L_2}$	$1.76 \times 10^4 \text{ min}^{-1}$	$\alpha_P$	$10.0 \text{ min}^{-1}$
$K_1$	$2.52 \times 10^{-2} (\mu\text{M})^{-2}$	$\beta_{L_1}$	$2.65 \times 10^3 \text{ min}^{-1}$
$K_{L_2}$	$9.7 \times 10^{-4} \text{ M}$		

**Table 12.2** The equations describing the evolution of the variables  $M$ ,  $B$ ,  $L$ ,  $A$ , and  $P$  in the Yildirim–Mackey model for the *lac* operon

$$\frac{d[M]}{dt} = \alpha_M \frac{1 + K_1[A]^n}{K + K_1[A]^n} + \Gamma_0 - \tilde{\gamma}_M[M]. \tag{12.25}$$

$$\frac{d[B]}{dt} = \alpha_B[M] - \tilde{\gamma}_B[B]. \tag{12.26}$$

$$\frac{d[L]}{dt} = \alpha_L \frac{[P][L_e]}{K_{L_e} + [L_e]} - \beta_{L_1} \frac{[P][L]}{K_{L_1} + [L]} - \beta_{L_2} \frac{[B][L]}{K_{L_2} + [L]} - \tilde{\gamma}_L[L]. \tag{12.27}$$

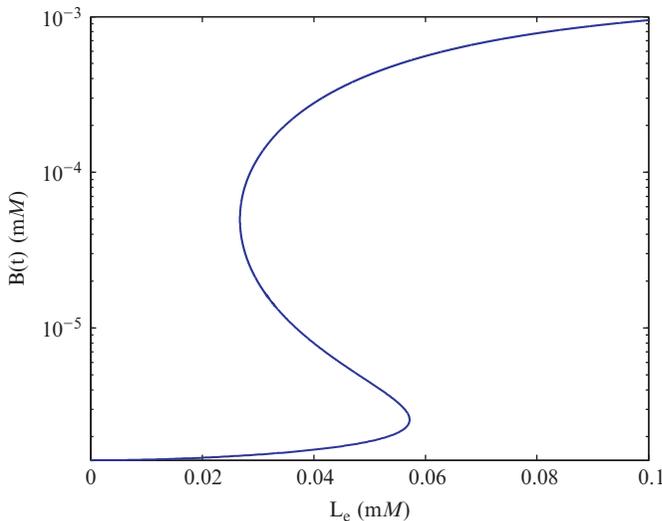
$$\frac{d[A]}{dt} = \alpha_A \frac{[B][L]}{K_L + [L]} - \beta_A \frac{[B][A]}{K_A + [A]} - \tilde{\gamma}_A[A]. \tag{12.28}$$

$$\frac{d[P]}{dt} = \alpha_P[M] - \tilde{\gamma}_P[P]. \tag{12.29}$$

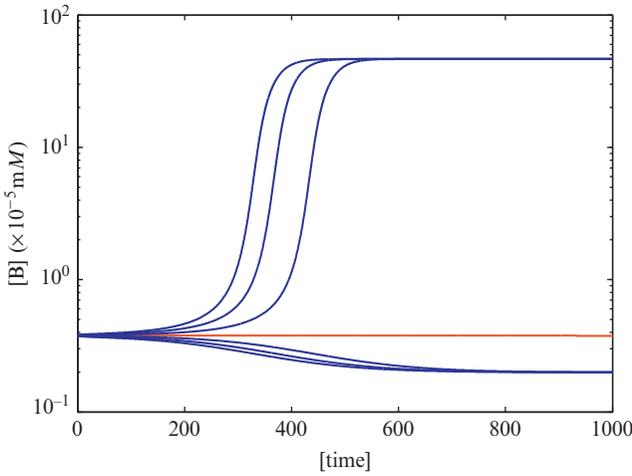
In this model  $\tilde{\gamma}_i = \gamma_i + \mu$ ,  $i \in \{M, B, L, A, P\}$ .

observed bistable behavior (Cohn and Horibata, 1959; Novick and Wiener, 1957; Ozbudak *et al.*, 2004), we set the left-hand side of each equation in the system Eqs. (12.25)–(12.29) to zero and solve the resultant system of five nonlinear equations for a range of  $L_e$  concentration after keeping all the other parameters as in Table 12.1 for  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$ . The result is shown in Fig. 12.10. Our modified model predicts that there is a physiologically range for the external lactose concentration that corresponds to the S-shaped curve in this figure. When the external lactose concentration falls in this range, the *lac* operon can have three coexisting steady states.

Figure 12.11 shows how the bistability arises in evolution of  $\beta$ -galactosidase concentration in the deterministic simulation of the model. In this simulation, all the parameters are kept constant as in Table 12.1 when  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$  and we chose  $[L_e]$  as  $[L_e] = 53 \times 10^{-3} \text{ mM}$ . As shown in Fig. 12.10, there are three steady states for this particular concentration of  $L_e$ . We calculate these steady state values numerically as in Table 12.3. To produce this figure, the initial values for the concentrations of all the proteins were kept constant at their steady state values on the middle branch of the S-shaped curve when  $[L_e] = 53 \times 10^{-3} \text{ mM}$  except *mRNA* concentration. Then three initial values of the *mRNA* concentration were chosen slightly below its steady value on the middle branch and another three initials were chosen slightly above its steady state



**Figure 12.10** Bistability arises in the *lac* operon model as the external lactose ( $L_e$ ) concentration changes when  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$ . Notice that the parameter values, there exists a range of  $L_e$  concentration for which there are three coexisting steady states for  $\beta$ -galactosidase concentration. Our calculations estimate this range as  $[0.026, 0.057] \text{ mM}$  of  $[L_e]$ .



**Figure 12.11** Semilog plot of  $\beta$ -galactosidase concentration over time showing effects of the initial values of concentration mRNA around the middle branch of S-shape curve in Fig. 12.10 in the numerical simulation.

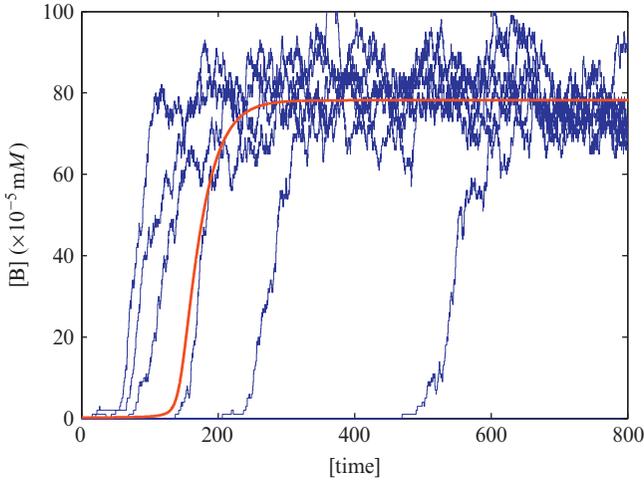
concentration on the same branch and the model equations in Eqs. (12.25)–(12.29) was solved numerically for  $[L_c] = 80 \times 10^{-3} \text{ mM}$ , which corresponds to the external lactose concentration on a steady on the upper branch of the S-shaped curve. When the simulation is started from an initial starting point that is exactly on the middle branch of S-shape curve, the  $\beta$ -galactosidase concentration stays constant over time (the horizontal line in Fig. 12.11) as it is a steady state for this system. Since the middle branch is unstable, small perturbations around the middle branch can kick the simulation either to the lower or the upper stable branches of S-shape curve. All the other runs converge to the stable steady states either on the lower branch or on the upper branch as seen in this simulation. We observe that the ones started initially above the steady state concentration of mRNA on the middle branch converged to the steady state on the upper branch, the ones started initially below the steady state concentration of mRNA on the middle branch converged to the steady state on the lower branch.

In Fig. 12.12, the deterministic and stochastic simulation of the Yildirim–Mackey *lac* operon model is shown. To produce this plot, we run six simulations by choosing the steady state value on the lower branch of the S-shaped curve as the initial starting point when  $[L_c] = 53 \times 10^{-3} \text{ mM}$  and  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$  while all other parameters are kept constant as in Table 12.1. As seen in this simulation, the average of the stochastic simulations are about the same as the solution of differential equations. Since we pick the initial concentrations from the bistable region, there is a slow transition before reaching to the steady state in both simulations.

**Table 12.3** The steady state values calculated from Eqs. (12.25)–(12.29) by setting the time derivatives zero

	$[M^*]$	$[B^*]$	$[A^*]$	$[L^*]$	$[P^*]$
Lower branch	$2.80 \times 10^{-6}$	$1.98 \times 10^{-6}$	$1.00 \times 10^{-2}$	$1.88 \times 10^{-1}$	$4.17 \times 10^{-5}$
Middle branch	$5.33 \times 10^{-6}$	$3.78 \times 10^{-6}$	$2.04 \times 10^{-2}$	$2.11 \times 10^{-1}$	$7.93 \times 10^{-5}$
Upper branch	$6.56 \times 10^{-4}$	$4.65 \times 10^{-4}$	$3.37 \times 10^{-1}$	$2.46 \times 10^{-1}$	$9.75 \times 10^{-3}$

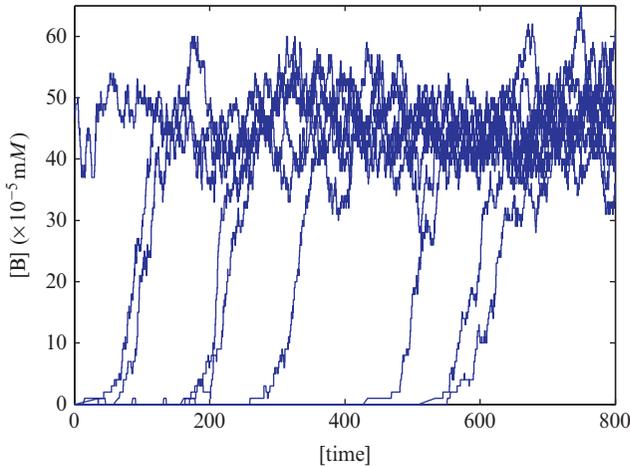
All the parameters are kept constant as in Table 12.1, when  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$  and  $[L_e] = 53 \times 10^{-3} \text{ mM}$  for which there exist three steady states (see Fig. 12.10).



**Figure 12.12** Deterministic and stochastic simulation of the Yildirim–Mackey *lac* operon model given by Eqs. (12.25)–(12.29). In this simulation, the solid lines show the ODE solutions and the broken lines represent the results of the stochastic simulations. To produce this plot, we chose the steady state value on the lower branch of the S-shaped curve as the initial value when  $[L_c] = 53 \times 10^{-3} \text{ mM}$  and  $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$  while all the other parameters are kept constant as in Table 12.1 and run six stochastic simulations for the external lactose concentration  $[L_c] = 80 \times 10^{-3} \text{ mM}$  which corresponds to a steady state value on the upper branch of the S-shaped curve in Fig. 12.10.

The deterministic model estimates this transition period about 120 min. The stochastic simulations predicts a significant variance in this transition period and estimate that this period may take up to 500 min for individual cells.

We investigate the effects of stochasticity in the bistable region. To this end, we run the stochastic simulation eight times starting from the stable steady state on the lower branch of the S-shaped curve and another eight runs starting from the stable steady state on the upper branch of the S-shaped curve for  $[L_c] = 53 \times 10^{-3} \text{ mM}$ . The results are shown in Fig. 12.13. In a bistable system, the random fluctuations can push the system from one stable steady state to the other one. The frequency of this transition is higher for systems with higher noise levels. We observe that all simulations starting from the lower branch of the S-shaped curve ended up converging to the stable steady state on the upper branch. However, simulations initialized at the upper branch never switch to the lower steady state and stay on the upper branch. This indicates that the steady state on the upper branch is more robust, and is resistive against fluctuations in the protein concentrations compared to the steady state on the lower branch. As seen in this simulation, the time required to shift from the lower steady state to the upper steady state can change



**Figure 12.13** Sixteen stochastic simulations of the Yildirim–Mackey *lac* operon model is shown, with the same parameters used for Fig. 12.12. Eight simulations use the lower steady state value as the initial condition, while the others use the upper steady state as the initial condition. We observe that all the simulations starting from the lower branch converge to the upper steady state and the simulations initialized from the upper branched stay on that steady state (only one of the simulations is plotted here).

significantly from one run to another. This transition can happen as early as in 60 min and as late as 600 min. Another surprising result is the variance at steady levels of  $\beta$ -galactosidase and lactose. When  $[L_c] = 53 \times 10^{-3}$  mM, the steady state concentrations of  $\beta$ -galactosidase and lactose concentration are around 50 and 23,000 mM, respectively. In general, we expected to see less variation when concentration of a molecular species is high. In other words, relative noise is less for high concentrations. However, our stochastic simulation results indicate that relative noise appears to be about the same for both of these proteins (results are not shown). One conclusion we can derive from this simulation result is about the sensitivity of concentration of  $\beta$ -galactosidase, that significant changes in the concentration of  $\beta$ -galactosidase is not likely to have an impact on the entire system, because it will most likely be dominated by noise anyway.

## 5. CONCLUSIONS AND DISCUSSION

Here, we present a brief introduction to mathematical modeling of regulatory biochemical reaction networks with some examples from enzyme kinetics and the couple systems that are capable of displaying oscillatory dynamics and bistable behaviors. We cover both deterministic and stochastic approaches and discuss the bistability and its origin from a mathematical point

of view in Section 2. We give the *lac* operon as a real-life example and show that this system is capable of bistable behavior for physiologically meaningful parameters ranges. We compare the stochastic and deterministic simulation results on the *lac* operon. All numerical computations in this study were performed in MatLab. There are software packages freely available to perform deterministic and stochastic simulation of biological and ecological networks (Adalsteinsson *et al.*, 2004; Kazanci, 2007).

We study the Yildirim–Mackey *lac* operon model using both deterministic and stochastic approaches and show that the model is capable of producing three coexisting steady states that correspond to the S-shaped curve in Fig. 12.10. The external lactose concentration for the bistability is estimated as  $[0.026, 0.057]$  mM of  $[L_e]$  which agrees well with the recent experimental result in (Ozbudak *et al.*, 2004).

In the bistable region, our stochastic simulation results indicate that the stable steady state on the lower branch of the S-shaped curve is less stable against noise than the steady state on the upper branch of the S-shaped curve. Furthermore, the fluctuations in the protein concentrations on the lower branch of the S-shaped curve are strong enough to shift the *lac* operon to the stable steady state on the upper branch (Fig. 12.13). Both the deterministic and stochastic simulations predicts that there is a significant transition period from bistable region ( $[L_e] = 53 \times 10^{-3}$  mM) to fully the induced state ( $[L_e] = 80 \times 10^{-3}$  mM). The deterministic model estimates this period about 2 h and the stochastic simulations predicts this period may take as long as 500 min for individual cells (Fig. 12.12).

To close, we would like to mention that both deterministic and stochastic methods have certain advantages and shortcomings. Deterministic simulations describe the average behavior, and are appropriate when the number of molecules in a system is large enough and molecules are spatially homogeneous. When the number of molecules is small, the stochastic methods simulate system behavior much better. Another major shortcoming of the deterministic simulation shows up in systems capable of multiple steady states. The deterministic solution of such a system always converges to a single stable steady state and stays there forever. However, in real life, constant switching behavior among steady states may happen due to inherent fluctuations within the system, as shown in Fig. 12.8. Dynamics such as this can only be captured by stochastic methods. On the other hand, the deterministic methods are often computationally more efficient and easier to implement.

In Section 3, we went over two systems that display significantly different behavior when simulated by deterministic and stochastic methods. Stochasticity may play crucial role in regulation of a dynamical system. There are many other biological systems observed with such properties, systems that display noise-induced stability (D’Odorico *et al.*, 2005) or stochastic resonance (Gammaitoni *et al.*, 1998).

## ACKNOWLEDGMENT

The work was partially supported by the New College of Florida Faculty Development Funds.

## REFERENCES

- Adalsteinsson, D., McMillen, D., and Elston, T. C. (2004). Biochemical network stochastic simulator (bionets): Software for stochastic modeling of biochemical networks. *BMC Bioinform.* **5**(24), 1–21.
- Beckwith, J. (1987). The lactose operon. In “*Escherichia coli* and Salmonella: Cellular and molecular biology,” (F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, and H. E. Umbarger, eds.), Vol. 2, pp. 1444–1452. American Society for Microbiology, Washington DC.
- Cohn, M., and Horibata, K. (1959). Inhibition by glucose of the induced synthesis of the  $\beta$ -galactosidase-enzyme system of *Escherichia coli*: Analysis of maintenance. *J. Bacteriol.* **78**, 613–623.
- D’Odorico, P., Laio, F., and Ridolfi, L. (2005). Noise-induced stability in dryland plant ecosystems. *Proc. Natl. Acad. Sci. USA* **102**(31), 10819.
- Gammaitoni, L., Hanggi, P., Jung, P., and Marchesoni, F. (1998). Stochastic resonance. *Rev. Mod. Phys.* **70**(1), 223–287.
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340.
- Gillespie, D. T. (2000). The chemical Langevin equation. *J. Chem. Phys.* **113**, 297.
- Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**, 1716.
- Gillespie, D. T., and Petzold, L. R. (2003). Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.* **119**, 8229.
- Kazanci, C. (2007). Econet: A new software for ecological modeling, simulation and network analysis. *Ecol. Modell.* **208**(1), 3–8, (Special Issue on Ecological Network Theory).
- Lewin, B. (2008). Genes. 9th edn. Jones and Bartlett Publishers, Sudbury, Massachusetts.
- Mackey, M., Santillán, M., and Yildirim, N. (2004). Modeling operon dynamics: The tryptophan and lactose operons as paradigms. *CR Biol.* **327**(3), 211–224.
- Novick, A., and Wiener, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA* **43**, 553–566.
- Ozbudak, E., Thattai, M., Lim, H., Shraiman, B., and Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740.
- Schleif, R. (2000). Regulation of the 1-arabinose operon of *Escherichia coli*. *Trends Genet.* **16**(12), 559–565.
- Scott, S. K. (1991). Chemical Chaos. 1st edn. Oxford University Press, Oxford and New York.
- Wilkie, J., and Wong, Y. M. (2008). Positivity preserving chemical Langevin equations. *Chem. Phys.* **353**(1–3), 132–138.
- Yildirim, N., and Mackey, M. (2003). Feedback regulation in the lactose operon: A mathematical modeling study and comparison with experimental data. *Biophys. J.* **84**(5), 2841–2851.
- Yildirim, N., Santillán, M., Horike, D., and Mackey, M. (2004). Dynamics and bistability in a reduced model of the *lac* operon. *Chaos* **14**(2), 279–292.