Chapter:

Computational Approaches for Modeling Intrinsic Noise and Delays in Genetic Regulatory Networks

Manuel Barrio[1], Kevin Burrage[2,3], Pamela Burrage[3], André Leier[4], Tatiana Márquez Lago[4]

[1]Departamento de Informática, University of Valladolid, Spain

mbarrio@infor.uva.es

[2]Comlab and OCISB,

The University of Oxford, Oxford, U.K

kevin.burrage@comlab.ox.ac.uk

[3]Institute for Molecular Bioscience,

The University of Queensland, Brisbane, Australia

{k.burrage,p.burrage }@imb.uq.edu.au

[4]Institute for Computational Sciences

ETH Zurich, Zurich, Switzerland

{leier,tmarquez}@inf.ethz.ch

Abstract

This chapter focuses on the interactions and roles between delays and intrinsic noise effects within cellular pathways and regulatory networks. We address these aspects by focusing on genetic regulatory networks that share a common network motif, namely the negative feedback loop, leading to oscillatory gene expression and protein levels. In this context, we discuss computational simulation algorithms for addressing the interplay of delays and noise within the signaling pathways based on biological data. We address implementational issues associated with efficiency and robustness. In a Molecular Biology setting we present two case studies of temporal models for the Hes1 gene (Monk, 2003; Hirata et al., 2002), known to act as a molecular clock, and the Her1/Her7 regulatory system controlling the periodic somite segmentation in vertebrate embryos (Giudicelli and Lewis, 2004; Horikawa et al., 2006).

1. Introduction

The mathematical modeling and simulation of genetic regulatory networks can provide insights into the complicated biological and chemical processes associated with genetic regulation. However, highly resolved computational models of such biochemical complexity can be very expensive and often infeasible and, thus, it is important that the models are kept simple but nevertheless capture the key processes.

Two vital aspects in modeling genetic regulatory networks are intrinsic noise and delays. Intrinsic noise arises in the system when there are small to moderate numbers of certain key molecules and is due to the uncertainty of knowing when a reaction occurs and which reaction it might be. Intrinsic noise is entirely different to extrinsic noise in which state changes are due to fluctuations in external conditions, such as temperature. These intrinsic noise effects can be modeled through the Stochastic Simulation Algorithm (SSA), first applied by Gillespie (1977) to simulate discrete chemical kinetics as the evolution of a discrete nonlinear Markov process.

Delays are intrinsic to slow biochemical processes that do not occur instantaneously and are often affected by spatial inhomogeneities. For instance, they are often associated with transcription and translation, two processes that imply other spatiotemporal processes often not explicitly modeled, such as (in eukaryotes) diffusion and translocation into and out of the nucleus, RNA polymerase activation, splicing, protein synthesis, and protein folding. These processes can take many minutes and so the effects are very important especially in the laying down of oscillating patterns of gene expression (Hirata et al., 2002). Monk (2003) notes that in mouse there is an average delay of 10–20 minutes between the action of a transcription factor on the promoter region of a gene and the appearance of the corresponding mRNA in the cytosol. Similarly, there is a delay of typically 1–3 minutes for the translation of a protein from mRNA.

By incorporating delays into the temporal model we can capture essential information on a macroscopic level, the delay can itself account for the multitude of biochemical processes and events on a microscopic time scale that render us unable to compute cell dynamics in real-time. Hence, we can expect more accurate and reliable predictions of cellular dynamics through the use of time delay models (Barrio et al., 2006).

One of the first people to consider feedback differential equation models for the regulation of enzyme synthesis was Goodwin (1965). An der Heiden (1979) then modified these ideas by including transport delays into Goodwin's model. The oscillatory behavior of

the ensuing delay differential equations (DDEs) as a function of the size of delays was investigated by an der Heiden. However, these DDE models act in the continuous deterministic regime and this regime is not always appropriate when considering small numbers of molecules such as in the case of genetic regulation with small numbers of transcription factors.

In a lovely set of experiments, Hirata et al. (2002) measured the production of *hes1* mRNA and Hes1 protein in mice. This work forms the basis of one of our case studies in Section 4.1. Serum treatments on cultured cells result in oscillations in expression levels for *hes1* mRNA and Hes1 protein in a two hour cycle with a phase lag of approximately 15 minutes between the oscillatory profiles of mRNA and protein. The oscillations in expression continue for 6 to 12 hours.

In order to explain the observed behaviors, Hirata et al. modified a mathematical model developed by Elowitz and Leibler (2000) for a synthetic gene network constructed in *E. coli* cells by introducing one gene from λ-phage. By postulating a Hes1 interacting factor as a third molecular species Hirata et al. obtained a system of three Ordinary Differential Equations (ODEs) that gives rise to sustained oscillatory behavior. However, there is no direct experimental evidence for such an interacting factor. Rather, the introduction of a third variable is due to the fact that certain systems of two ODEs cannot generate sustained oscillations. This observation together with the experimental results of Hirata et al. led to a number of papers in which simple coupled delay differential equations were developed in order to explain the sustained oscillations without recourse to the addition of a third variable (Monk, 2003; Jensen et al., 2003; Lewis, 2003; Bernard et al., 2006).

Barrio et al. (2006) took a different approach from the above authors and tried to explain the results of Hirata et al. by taking proper account of both time delays and intrinsic randomness. They developed a Delay Stochastic Simulation Algorithm (DSSA) that generalizes the Stochastic Simulation Algorithm (SSA) to the delayed setting. Independently, Bratsun et al. (2005) developed a delay SSA without considering waiting times for delayed reactions while only non-consuming reactions can be specified to be delayed. More recently, Cai (2007) introduced a direct delay SSA method and showed that both, the DSSA by Barrio et al. and the direct method are exact stochastic simulation algorithms for chemical reaction systems with delays. The experimental results of Hirata et al. seemed to be better explained through the delay stochastic simulation algorithm approach rather than through delay differential equations (Barrio et al., 2006).

When modeling biological systems with large numbers of molecules and/or rate constants, the time steps in stochastic simulation algorithms can become very small and, hence, the simulation can be computationally highly expensive. By consequence, this limits the feasible 'real-time' span of the simulations. In order to reduce the computational load we need new algorithms that still model intrinsic noise in a delayed setting but overcome the issues of small step sizes. Temporal coarse-graining has been considered through the use of τ-leap methods (Gillespie, 2001; Tian and Burrage, 2004; Peng et al. 2007, Anderson, 2007, 2008), and similar ideas have been applied in the delay setting (Leier et al., 2008(a)), thus rendering an efficient algorithm that yields accurate simulations in time spans that are long enough to be of actual interest to the experimentalists.

Lastly, temporal delay models lack spatial resolution but nevertheless allow for portraying spatial aspects of cellular processes by compartmentalization, that is, by distinguishing between identical molecular species according to their location. Recent research suggests that molecular translocation processes can be well captured and modeled by means of time delayed processes with specific delay distributions. However, it is worth

mentioning that spatial algorithms are not replaceable in all cases. Examples of the latter are scenarios with high spatial heterogeneity, anisotropies, or when single-particle tracking becomes strictly necessary. Spatial highly-resolved algorithms are computationally most expensive, and coarse-graining techniques have also been developed for this case (Chatterjee and Vlachos, 2005; Chatterjee and Vlachos, 2006).

The outline of this Chapter is as follows. In section 2 we give an overview of some of the approaches to the temporal modeling of chemical kinetics. In section 3 we present various types of simulation algorithms with and without delays and discuss how we can improve the accuracy and robustness by so-called τ leap approaches. Section 4 gives two case studies: the Hes1 molecular clock and the Her1/7 complex which plays a role in somite formation in zebrafish. Section 5 presents some conclusions.

2. Modeling Chemical Kinetics

Modeling and simulations are valuable tools for investigating complex biochemical systems. Not only do they allow us to determine if a proposed reaction mechanism is consistent with observed experimental results, but they can also aid experimental design techniques by exploring reaction network interactions with relative ease. The choice for a particular modeling approach depends on several factors, such as molecular concentrations, distributions, the type of reactions and their time scales, whether discreteness and internal noise have noticeable macroscopic effects and, lastly, if the model requires spatial information.

Deterministic models assume a time evolution that is both continuous and predictable. However, randomness is intrinsic to biological systems, where system behavior is typically represented by noisy signals. Often the most important source of stochasticity stems from the fact that molecular reactions are random events, as it is impossible to say with certainty the specific type of reaction that will happen next, or when or where such event is to occur. Moreover, low molecular concentrations, coupled to random diffusion, are an important source of spatial inhomogeneity and stochastic variation.

In a purely temporal setting, and when there are large numbers of molecules present, chemical reactions are modeled by ordinary differential equations that are based on the laws of Mass Action and the fact that reaction rates can be estimated on the basis of average values of the reactant density. Any set of *m* chemical reactions can be characterized by two sets of quantities: the stoichiometric vectors (update rules for each reaction) $\nu_1,...,\nu_m$ and the propensity functions $a_1(X(t)),...,a_m(X(t))$. The propensity functions represent the relative probabilities of each of the *m* reactions occurring. Here X(t) is the vector of concentrations at time *t* of the *N* species involved in the reactions. The ODE that describes this chemical system, under the Law of Mass Action, is given by

$$X'(t) = \sum_{j=1}^{m} \nu_j a_j(X(t)). \quad (1)$$

In order to make this clearer we give a simple example for Michaelis–Menten kinetics. This system involves a substrate (S), an enzyme (E), a complex (C) and a product (P). The kinetics can be written as

$$E + S \longrightarrow C,$$
$$C \longrightarrow E + S,$$

$$C \longrightarrow E + P.$$

Let X(t) be the concentration of (E(t), S(t), C(t)) then the stoichiometric vectors (or the update rules for each of the three reactions) are

$$\nu_1 = (-1,-1,1)^T, \quad \upsilon_2 = (1,1,-1)^T, \quad \nu_3 = (1,0,-1)^T.$$

The time dependent propensity functions $a_1(X(t)),...,a_m(X(t))$ are the relative probabilities of each of the three reactions occurring, respectively, and are given by

$$a_1(X) = k_1 ES$$
$$a_2(X) = k_2 C$$
$$a_3(X) = k_3 C.$$

In this case (1) becomes

$$X_1^{'} = -k_1 X_1 X_2 + (k_2 + k_3) X_3$$
$$X_2^{'} = -k_1 X_1 X_2 + k_2 X_3$$
$$X_3^{'} = k_1 X_1 X_2 - (k_2 + k_3) X_3.$$

Often in such systems there is a conservation of molecular numbers (here $X_1'+X_3'=0$) and so one or more equations can be removed. Additional equations can be removed by the use of the Quasi-Steady State Assumption (QSSA). Under the QSSA it is assumed that the fast reactions go to equilibrium much more quickly that the slow reactions. Thus a system of algebraic equations can be solved at the "fast equilibrium" and this solution substituted back into the original system, thus reducing the dimension and altering the propensity functions to include nonlinear Hill functions.

In the case of small numbers of molecules the appropriate modeling formulation is the Stochastic Simulation Algorithm, as ODEs can only describe a mean behavior. The SSA is essentially an exact procedure that describes the evolution of a discrete nonlinear Markov process. It accounts for the inherent stochasticity (internal noise) of the *m* reacting channels and only assigns integer numbers of molecules to the state vector. At each step, the SSA simulates two random numbers (representing probabilities) from the uniform distribution U[0,1] to evaluate an exponential waiting time, τ, for the next reaction to occur and an integer *j* between 1 and *m* that indicates which reaction occurs. The state vector is updated at the new time point by the addition of the j^{th} stoichiometric vector to the previous value of the state vector, that is

$$X(t+\tau) = X(t) + \nu_j.$$

The main limiting feature of SSA is that the time step can become very small, especially if there are large numbers of molecules or widely varying rate constants. In order to overcome these limitations, a number of different approaches (so called τ-leap methods) have been suggested in which the sampling of likely reactions is taken from either Poisson (Gillespie, 2001) or Binomial (Tian and Burrage, 2004) distributions. In these cases a much larger time step can be used at the loss of a small amount of accuracy. Cao et al. (2006) have analyzed effective strategies for choosing the step size in τ-leap methods. The reason sampling occurs from a Poisson distribution is due to the fact that the SSA can also be viewed as a type of τ leap method based on Poisson sampling (Kurtz, 1971). On the other hand, Binomial sampling is valid because as the number of molecules becomes large, Poisson random variables are well approximated by Binomial random variables.

A very different approach is to note that the discrete nonlinear Markov process described by the SSA has a probability density functions that is the solution of the so-called Chemical Master Equation (CME). The CME is a discrete parabolic partial differential equation in which there is an equation for each configuration of the State Space. When the State Space is enumerated, the CME becomes a linear ODE and the probability density function takes the form

$$p(t) = e^{At} p(0)$$

where *A* is the state-space matrix. Even for relatively small systems, the dimension of *A* can be in the millions, so it would appear that this is not a computationally feasible approach. However, one should consider that not all of the states are reachable. Furthermore, a proposed finite state projection algorithm (Munsky and Khammash, 2006) reduces the size of the matrix *A*. Then one can use Krylov subspace techniques (Burrage et al., 2006) to efficiently compute the exponential of a matrix times a vector, making the computation of the probability density function directly a very feasible technique (MacNamara et al., 2007).

Finally, it is important to note that there is a regime intermediate to the discrete stochastic regime and the continuous deterministic ODE regime in which the internal noise effects are still significant but continuity arguments can apply. This leads to the so-called Chemical Langevin Equation (CLE) that is an Itô stochastic ordinary differential equation (SDE), driven by a set of Wiener processes that describes the fluctuation in the concentrations of the molecular species. The CLE preserves the correct dynamics for the first two moments of the SSA and takes the form

$$dX = \sum_{j=1}^{m} \nu_j a_j(X(t)) + B(X(t))dW(t).$$

Here $W(t) = (W_1(t),...,W_N(t))$ is a vector of N independent Wiener processes whose increments $\Delta W_j = W_j(t+h) - W_j(t)$ are $N(0,h)$ and where

$$B(x) = \sqrt{C}, \quad C = (\nu_1,...,\nu_m)Diag(a_1(X),...,a_m(X))(\nu_1,...,\nu_m)^T.$$

Here *h* is the time discretization step. This formulation can be derived from the Poisson formulation of the SSA by noting that as $Th \to \infty$ with $h \to 0$,

$$\begin{aligned} P(Th) \to N(Th,Th) &= Th + \sqrt{T}h\,N(0,1) \\ &= Th + \sqrt{T}\,\Delta W. \end{aligned}$$

Effective numerical methods designed for the numerical solution of SDEs (such as the Euler-Maruyama method) can be used to simulate the chemical kinetics in this intermediate regime. Furthermore, adaptive multiscale methods have been developed which attempt to move back and forth between these three regimes as the numbers of molecules change (Burrage et al., 2004).

None of these frameworks explicitly incorporate delay affects but in fact the same modeling regimes arise in a natural fashion if delay is included. These have been thoroughly explored in Barrio et al. (2006) and Tian et al. (2007) in terms of the same modeling regimes mentioned above. We now discuss some of the issues when incorporating noise and delays.

3. Simulation algorithms

In recent years, discrete stochastic simulation techniques have been widely used to help understand the dynamic behavior of biochemical systems such as genetic regulatory networks and intra-cellular and inter-cellular signaling pathways when there are small to

moderate numbers of molecular species involved. In addition to the methods mentioned above, other simulation type methods have also been proposed recently, for example, Gibson and Bruck's next reaction method (2000), Gillespie's continuous model (2000) and the probability-weighted Monte-Carlo approach by Resat et al. (2001). In this section we review some of these approaches without and with delays and then discuss extensions via tau leaping strategies which can dramatically improve robustness and computational performance.

3.1 SSA

The SSA (Stochastic Simulation Algorithm) is a numerical Monte Carlo procedure that can be used to simulate the time evolution of a set of molecular species affected by a given set of reactions. It was introduced by Gillespie (1977) as an exact calculation that generates simulated trajectories of the system state. These trajectories are numerical realizations of the Chemical Master Equation (CME). It is important to note that the SSA is based on a fundamental stochastic premise that defines the probability, given a particular state that one reaction will occur in the next infinitesimal time internal. This assumption is used without approximation by the SSA and makes it exact with respect to the CME.

More precisely, consider a well-stirred volume Ω of molecules containing N molecular species $\{S_1,...,S_N\}$ that interact at constant temperature through M chemical reactions $\{R_1,...,R_M\}$. Given the system state at a particular time $\mathbf{X}(t)$ which represents the number of molecules of each species, we can define for each reaction R_j (j=1,...,M) its propensity function $a_j(\mathbf{x})$ in a given state $\mathbf{X}(t)=\mathbf{x}$ so that

> $a_j(\mathbf{x})dt$ = probability that one R_j reaction will occur somewhere inside Ω in the next infinitesimal time interval [t,t+dt).

Additionally, each reaction is characterized by its stoichiometric vector ν_j that defines the state change in the number of species due to reaction R_j.

The procedure to generate simulated trajectories of $\mathbf{X}(t)$ is based on the probability function of the two random variables: (1) the time τ to the next occurring reaction, and (2) the index j of the next reaction. Given a current state **x**, the probability of state change per unit of time is constant ($a_0(\mathbf{x})$) and so the waiting time to the next reaction is an exponential random variable with mean $1/a_0(\mathbf{x})$. The reaction index j is an integer random variable with point probabilities

$$a_j(\mathbf{x}) / a_0(\mathbf{x}), \text{ where } a_0(\mathbf{x}) = \sum_{k=1}^{M} a_k(\mathbf{x}).$$

These two random variables and their distributions are the basis of the SSA. One of the simplest Monte Carlo procedures for generating time and index of the next reaction is the so-called 'direct method'. Two independent random numbers r_1 and r_2 are drawn from the uniform distribution in the unit interval U(0,1), and then τ is assigned as

$$\tau = \frac{1}{a_0(x)} \ln(1/r_1),$$

while j is the reaction index that satisfies

$$\sum_{k=1}^{j-1} a_k(x) < r_2 \cdot a_0(x) \le \sum_{k=j}^{j} a_k(x).$$

Then the system is updated by $\mathbf{x}(t+\tau) = \mathbf{x}(t) + \nu_j$, and the procedure is repeated to evolve the system through time. The following is an algorithmic representation of the direct method.

Algorithm 1: SSA

Data: reactions defined by reactant and product vectors, stoichiometry, reaction rates, initial state $X(0)$, simulation time T
Result: state dynamics
begin
  while $t < T$ **do**
    generate U_1 and U_2 as $U(0,1)$ random variables
    $a_0(X(t)) = \sum_{j=1}^{m} a_j(X(t))$
    $\theta = \frac{1}{a_0(X(t))} \ln(1/U_1)$
    select j such that
      $\sum_{k=1}^{j-1} a_k(X(t)) < U_2 a_0(X(t)) \leq \sum_{k=1}^{j} a_k(X(t))$
    $X(t+\theta) = X(t) + \nu_j$
    $t = t + \theta$
end

3.2 Delay SSA

Biological processes often involve complex reactions and mechanisms that cannot be considered instantaneous. Reactants are processed and products are not present until a certain future time point. This time delay should be incorporated into our computational models if we want to capture a faithful representation of the biological process. Additionally, delays are often important parameters that affect the dynamic evolution of the system. A system of DDEs can take the general form

$$y' = f(t, y(t), y(t-\tau)),$$

and in the case of chemical kinetics as described by (1), the DDE formulation is

$$X'(t) = \sum_{j=1}^{m} \upsilon_j a_j\left(X\left(t-\tau_j\right)\right).$$

There are a number of suitable numerical methods for solving such systems, some of which are implemented in MATLAB. However, if intrinsic noise is important then we need a generalization of the *stochastic simulation algorithm* (SSA) for chemical kinetics with delayed reactions. The DSSA differs from the SSA by making a clear distinction between the reaction waiting time and reaction delay. The former is the time between two consecutive reactions whereas the latter is the time elapsed from the processing of the reactants to the appearance of the products.

Simulation proceeds in the standard way (SSA) if non-delayed reactions take place. However, if the next reaction index points to a delayed reaction then we have to distinguish between two different types: consuming and non-consuming. In case of non-consuming reactions, the corresponding reactants and products are not updated. Instead, the state update is scheduled for 'present time + delay' which will be reached in a future simulation step. When that happens, the last drawn reaction is ignored and instead the state is updated according to the delayed reaction. Simulation continues at the delayed reaction time point. On

the other hand, if the reaction is consuming, reactants and products of delayed consuming reactions must be updated separately: (1) reactant consumption updates the state when the delayed reaction is selected and (2) product generation is updated when the reaction is completed.

The trajectories simulated by SSA are numerical realizations of the state evolution **X**(t). Additionally, the probability density function of **X**(t) is completely determined by the Chemical Master Equation. Similarly, a CME for the DSSA, namely a DCME, has been derived from first principles and the DSSA has a corresponding representation as a system of delay differential equations (DDEs) – see Barrio et al. (2006) and Tian et al. (2008).

Algorithm 2 is an algorithmic description of the DSSA dealing with both delayed and non-delayed, as well as with consuming and non-consuming, reactions. Time steps are defined either by a next reaction waiting time or by a delayed time update.

3.3 Spatial Methods

In many Cell Biology settings spatially resolved simulations are mandatory. Some common examples in which spatial simulations are unavoidable are systems embedded in complex spatial structures, molecular motion described by low diffusion rates, or systems containing significantly low numbers of molecules, to name a few. The most straightforward spatial technique is through reaction-diffusion partial differential equations. However, this approach is only valid if dealing with large molecular concentrations and when noise is not amplified throughout the system. If at least one of these conditions fails to hold, one must rely on spatial stochastic simulators, which can be discrete or continuous in nature and have different levels of spatial resolution.

It should always be kept in mind that there is a trade-off between simulation time and resolution. That is, the more highly-resolved, the more computationally expensive these simulations become. The highly resolved end of the discrete spatial stochastic simulators spectrum is represented by lattice and off-lattice particle based methods. In lattice methods a two-dimensional or three-dimensional computational lattice is used to represent a membrane or the interior of some part of a cell (Turner et al., 2004; Morton-Firth and Bray, 1998; Nicolau et al., 2006). Such a lattice is then "populated" with particles of different molecular species that may diffuse throughout the simulation domain by jumping to empty neighboring sites and, depending on user-specified reaction rules, interacting chemically with a certain probability. Such lattice-based simulators are commonly referred to as Kinetic Monte Carlo Methods.

In off-lattice methods, particles have their own specific spatial coordinates and reaction bins whose size depends on the particular diffusion rates are drawn around them. If one or more molecules happen to be inside such a bin, appropriate chemical reactions can take place with a certain probability, and if a reaction is readily performed, the reactant particles are flagged. It should be noted that in off-lattice methods, the domains and/or compartments are usually still discretized to efficiently localize particles.

Particle methods can provide very detailed simulations of highly complex systems at the cost of exceedingly large amounts of computational time and, possibly, restrictions on the size of the simulation domain. Hence, such detailed simulations can often only yield short simulation time spans that may not be of sufficient interest to experimentalists.

Algorithm 2: DSSA

Data: reactions defined by reactant and product vectors, consuming delayed reactions are marked, stoichiometry, reaction rates, initial state $X(0)$, simulation time T, delays

Result: state dynamics

begin

  while $t < T$ **do**

    generate U_1 and U_2 as $U(0,1)$ random variables

    $a_0(X(t)) = \sum_{j=1}^{m} a_j(X(t))$

    $\theta = \frac{1}{a_0(X(t))} \ln(1/U_1)$

    select j such that

      $\sum_{k=1}^{j-1} a_k(X(t)) < U_2 a_0(X(t)) \leq \sum_{k=1}^{j} a_k(X(t))$

    if *delayed reactions are scheduled within* $(t, t+\theta]$ **then**

      let k *be the delayed reaction scheduled next at time* $t+\tau$

      if k *is a consuming delayed reaction* **then**

        $X(t+\tau) = X(t) + \nu_k^p$ (update products only)

      else

        $X(t+\tau) = X(t) + \nu_k$

      $t = t + \tau$

    else

      if j *is not a delayed reaction* **then**

        $X(t+\theta) = X(t) + \nu_j$

      else

        record time $t + \theta + \tau_j$ for delayed reaction j with delay τ_j

        if j *is a consuming delayed reaction* **then**

          $X(t+\theta) = X(t) + \nu_j^s$ (update reactants)

      $t = t + \theta$

end

3.4 Coarse Grained Methods

A major drawback of delayed and non-delayed, spatial and non-spatial stochastic simulation algorithms are their high computational costs when dealing with large numbers of molecules or widely varying rate constants. These factors inevitably result in exceedingly small simulation time steps, making the overall simulation computationally expensive or even infeasible. In order to reduce the computational load, we can coarsen the simulation, accounting for many events in one single larger time step. This is the general idea behind the so-called τ-leap methods, where the simulation advances in time leaps while updating the system state according to a reasonably good approximation for the accumulated number of reactions (and diffusions if a spatial simulation) within the time step.

3.4.1 τ-leap Methods

Gillespie (2001) proposed the *Poisson τ-leap method* in which the number of reactions in each τ-leap are sampled from a Poisson distribution, and the τ step is controlled by a selection strategy that depends on a pre-specified control parameter ε, such that $0<\varepsilon<<1$.

The update procedure for the Poisson τ-leap method can be written as $x(t+\tau) = x(t) + \sum_{j=1}^{M} K_j \nu_j$, where $K_j = P(a_j(X(t))\tau)$, for reactions $j = 1,\ldots,M$, is a sample from the Poisson distribution with mean $a_j(X(t))\tau$. Further improvements were made by Gillespie and Petzold (2003), Rathinam *et al.* (2003), and Cao *et al.* (2005, 2006).

However, samples from a Poisson distribution range from zero to unbounded values. Hence, when updating the system, negative numbers of molecules can occur if larger step sizes are used. In order to avoid this, Tian and Burrage (2004) and later Chatterjee *et al.* (2006) proposed the Binomial τ-leap method where the numbers of reactions in a leap are drawn from a Binomial distribution. Thus, the various K_j take the form $K_j = B(N_j, P_j)$, where there are some subtleties in the form of the N_j and P_j, and such variables N_j and P_j represent the sample size and probability of occurrence of reaction type *j*, respectively. Auger et al. (2006) presented a modification to the original Binomial τ-leap method which is a more robust implementation than the original formulation. Furthermore, Anderson (2007, 2008) has shown interesting connections between sampling from the Poisson and Binomial distributions in the context of τ-leap methods in both a non-delayed and delayed setting.

Recently, Peng *et al* (2007) developed a modified Binomial τ-leap method that estimates the number of reaction products within a τ-leap step allowing them to participate in additional reactions in the same leap. However, Leier et al. (2008) show that such an approach may not accurately describe complex dynamics including time delays, and they propose a generalized τ-leap method, that is described in more detail in Section 3.4.2. Lastly, τ-leap methods can also be extended to the spatially resolved spectrum, where the simulation advances in time leaps that account for several molecular diffusion and reaction events, as shown by Marquez-Lago and Burrage (2007) and described in Section 3.4.3.

3.4.2 Bτ-DSSA

Initial Binomial τ-leap algorithms (Tian and Burrage, 2004; Peng et al., 2007) were not able to capture accurately the dynamics of certain chemical kinetics compared to the exact SSA/DSSA approach, due to insufficient numbers of reactions drawn in τ-leap steps. In Leier et al. (2008) a new generalized Binomial τ-leap method (Bτ-DSSA) is presented that addresses the difficulties associated with complex chemical kinetics and introduces delays into the Binomial τ-leap framework. A description of the Bτ-DSSA is given in Algorithm 3.

Estimating a proper maximum number N_j of potential reaction events of type R_j for the Binomial random variables $B(N_j, P_j)$ is crucial for an accurate reproduction of system dynamics. For specific reactions, Table I shows how to calculate N_j assuming R_j is an isolated reaction that does not share reactants with any other reactions. While this estimation is straightforward for isolated, elementary reaction, it is less obvious for chemical kinetics involving large, interacting reaction networks where multiple reactions share the same reactants.

Reaction R_j	Propensity a_j	N_j	Stoichiometric Coefficients
1st order $S_k \xrightarrow{c_j} S_l$	$a_j = c_j \cdot X_k$	X_k	$\nu_{j,k} = -1,$ $\nu_{j,l} = 1$
Heterodimeric $S_k + S_l \xrightarrow{c_j} S_m$	$a_j = c_j \cdot X_k X_l$	$\min\{X_k, X_l\}$	$\nu_{j,k} = \nu_{j,l} = -1,$ $\nu_{j,m} = 1$
Homodimeric $S_k + S_k \xrightarrow{c_j} S_l$	$a_j = c_j \cdot X_k (X_k - 1)/2$	$\left\lfloor \frac{X_k}{2} \right\rfloor$	$\nu_{j,k} = -2,$ $\nu_{j,l} = 1$
Hill type $S_k \xrightarrow{c_j f} S_k + S_l$	$a_j = c_j \cdot f(X_k)$ where $f(X_k(t)) = 1 - \frac{1}{1 + (X_k(t)/X_0)^h}$ (activation) or $f(X_k(t)) = \frac{1}{1 + (X_k(t)/X_0)^h}$ (inhibition) with Hill coefficient *h*.	constant, $N_j >> 1$	$\nu_{j,l} = 1$

Table I: Some simple reactions R_j and their corresponding propensities a_j, stoichiometric coefficients $\nu_{j,\cdot}$ and maximum number of potential reaction events N_j. Hill functions are often used to describe the regulatory effect of one or more transcription factors on the chemical kinetics. For a Hill function depending on a single transcription factor X_k this results in the propensity $a_j = c_j \cdot f(X_k)$. Calculating the $N_j(x)$ for Hill-type reactions involves some subtlety. For Hill type reactions, Leier et al. (2008) define $N_j(x) = C$ where C is some constant. Simulations show that, unless C is too small (< 10), it has no noticeable effect on the simulation outcome.

The Bτ-DSSA samples reaction numbers from Binomial distributions $B(N_j'', P_j)$ (Step 5 in Algorithm 3). Here, $N_j'' = N_j(x, \xi)$, with $\xi \equiv (\xi_1, \ldots, \xi_M)$ and $\xi_i \le N_i(x)$, is the maximal number of potential reaction events of type R_j when $\xi_1, \ldots, \xi_M$ reactions of $R_1, \ldots, R_M$ occur in the τ-step. For $N_j(x, \xi)$ it is assumed that $\xi_j, \ldots, \xi_M = 0$ since only the already sampled reaction numbers $\xi_1, \ldots, \xi_{j-1}$ are considered. However, unlike the original Binomial τ-leap method by Tian and Burrage (2004), the N_j are calculated considering only those reactions R_i (and hence ξ_i) that share reactant species with R_j. Figure 1 illustrates the difference.

As a consequence, in the Bτ-DSSA the maximal number of potential reaction events is usually larger than in the original Binomial τ-leap method. Numbers of delayed reactions are sampled in the same way as numbers of non-delayed reactions. The update of the system state (Step 6 in Algorithm 3) has to distinguish between delayed consuming and non-consuming reactions scheduled within the τ-leap, but also has to sample the update times of all delayed reactions drawn for the τ-leap.

Algorithm 3: Bτ-DSSA

Data: reactions defined by reactant and product vectors, consuming delayed reactions are marked, stoichiometry: $\nu = -\rho + \pi$ (with ρ and π being update vectors for left-hand-side and right-hand-side of reaction, respectively), reaction rates, initial state $X(0)$, simulation time T, delays, pre-specified $K \in [1, 10]$

Result: state dynamics

begin

1. Calculate $a_1(x), \cdots, a_m(x)$, $a_0(x) = \sum_{j=1}^{m} a_j(x)$; and $N_1(x), \cdots, N_m(x)$

2. Choose a τ-selection procedure: update corresponding variables

3. Check step-size conditions. For each reaction R_j
Calculate $N'_j = \min\{N_i(x), i \in I_j\}$ and $a'_j(x) = \sum_{i \in I_j} a_i(x)$
If $(N'_j > 0)$ AND $(a'_j(x)\tau/N'_j > 1)$ then $\tau = N'_j/a'_j(x)$

4. If $\tau \leq K/a_0$ perform a normal (D)SSA step, otherwise go to (5)

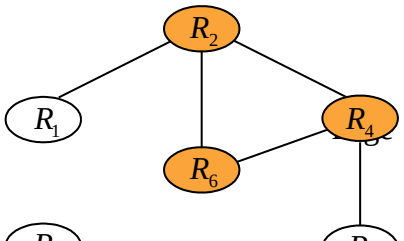
5. Sample. Initialise $\xi \equiv (\xi_1, \cdots, \xi_m) = 0$
For each reaction R_j
Calculate $N''_j = N_j(x, \xi)$
Generate a sample value ξ_j for the reaction number of type R_j
If $N''_j > 0$ then $\xi_j = B(N''_j, P_j)$ with $P_j = a_j(x)\tau/N''_j$ else $\xi_j = 0$.

6. Update non-delayed and delayed reactions. The subscripts nd_j and d_j represent the j^{th} non-delayed and delayed reaction, respectively. M' denotes the number of non-delayed reactions.
6.1 Delayed R_{d_j} with delay δ_j:
Record ξ_{d_j} random update time points $t + \delta_j + u_k\tau$ with $u_k \in U(0,1), k = 1, \cdots, \xi_{d_j}$
If R_{d_j} is a consuming, delayed reaction,
update $x(t+\tau) = x(t+\tau) + \rho_{d_j}$
6.2 Non-delayed: $x(t+\tau) = x(t) + \sum_{j=1}^{M'} \xi_{nd_j}\nu_{nd_j}$
6.3 Delayed (scheduled within $[t, t+\tau)$) :
$x(t+\tau) = x(t+\tau) + \pi_{d_j}$ (consuming, delayed reactions)
$x(t+\tau) = x(t+\tau) + \nu_{d_j}$ (remaining reactions)

end

Numerical simulations reveal that, unlike previous Binomial τ-leap methods, the Bτ-DSSA is better able to accurately capture the dynamics of oscillating patterns of gene expression. In such systems delayed reactions play a crucial role in maintaining the cyclic behavior and sampling too many or an insufficient number of delayed reactions will inevitably lead to a different cycle frequency. For the applications in Section 4.2, the Bτ-DSSA was able to reproduce the oscillatory dynamics both accurately and significantly faster than the DSSA. In case of the Her1/7-model for 5 coupled cells, Bτ-DSSA was 70 to 100 times faster than the DSSA implementation of Barrio et al. (2006).

R_1: $A \longrightarrow B$,



$$R_2: \quad A + B \longrightarrow C,$$
$$R_3: \quad C \longrightarrow A + B,$$
$$R_4: \quad B + C \longrightarrow D,$$
$$R_5: \quad D \longrightarrow B + C,$$
$$R_6: \quad B + E \longrightarrow F,$$

Figure 1: Artificial chemical kinetics system. The set of reactions R_1 to R_6 constitutes a network where two reactions, i.e. two vertices, are connected by an edge if and only if they have one or more common reactant species. The network has two connected subnetworks, $\{R_5\}$ and $\{R_1, R_2, R_3, R_4, R_6\}$. In the original Binomial τ-leap formulation, the maximum number of potential reaction events of type R_6 was calculated as the minimum N_i (see Table I) over the subnetwork $\{R_1, R_2, R_3, R_4, R_6\}$ (the subnetwork that R_1 belongs to). The Bτ-DSSA calculates $N_6(x, \xi)$ considering only R_6 and its direct (shaded) neighbors: $N_6(x, -, \xi_2, -, \xi_4, -, 0) = \min\{x_2 - \xi_2 - \xi_4, x_5\}$ with $x_2 = [B]$ and $x_5 = [E]$.

3.4.3 Bτ-SSSA

As mentioned before, particle methods can provide very detailed simulations at the cost of exceedingly large amounts of computational time and, possibly, restrictions on the size of the simulation domain. In other words, we may need to coarsen the simulation in order to provide a spatially resolved method that yields accurate chemical kinetics in meaningful simulation times that are of actual biological interest to experimentalists.

The idea behind τ-leaping in space is to account for several diffusion and reaction events in one larger time step, without compromising spatial nor temporal accuracy. Marquez-Lago and Burrage (2007) presented the Binomial τ-leap Spatial Simulation Algorithm, Bτ-SSSA, a coarse-grained version of an existing spatial stochastic simulation algorithm known as the next subvolume method (Elf and Ehrenberg, 2004; Elf et al., 2003; Hattne et al. 2005).

The next subvolume method is a generalization of the SSA, where the volume is divided into separate subvolumes that are small enough to be considered homogeneous by diffusion over the time scale of the reaction. At each step, the state of the system is updated by performing an appropriate reaction or by allowing a molecule to jump at random to a neighboring subvolume, where diffusion is modeled as a unary reaction with rate proportional to the two dimensional molecular diffusion coefficient divided by the length of a side of the subvolume. In this way, diffusion inside the algorithm becomes another possible event with a propensity function and follows the same update procedure as chemical reaction. Then, the expected time for the next event in a subvolume is calculated similarly to the SSA, including the reaction and diffusion propensities of all molecules contained in that particular subvolume at that particular time. However, time for next events will only be recalculated for those SVs that were involved in the current time step, and they are re-ordered in an event queue.

A natural extension of the next subvolume method is to perform τ-leaps that account for one or more diffusion and reaction events, the idea behind Bτ-SSSA (Marquez-Lago and Burrage, 2007). At each iteration, the subvolume with shortest reaction-diffusion τ-leap is selected, which is to be found at the top of the time event queue. Then, all randomly chosen but possible events inside such subvolume are executed, a new τ-leap for all subvolumes that were involved in the current τ-leap is calculated, the time event queue in increasing time is

reordered, and the subvolume indicated by the top of the time event queue is chosen. The algorithm is complicated and the reader can refer to the description in the article.

4. Case Studies

In this section we present results from two studies involving Notch signaling molecules. The first model is a model of hes1 auto-inhibition by Hes1 proteins in mouse (Monk, 2003; Barrio et al. 2006). The second model (Figure 2) describes the Delta-Notch dependent synchronization of Her1 and Her7 protein levels in a 1-dimensional array of cells in zebrafish (Lewis, 2003; Horikawa et al., 2006). In this model, the two linked genes her1 and her7 are autorepressed by their own gene products and positively regulated by Delta-Notch cell-cell signaling that leads to oscillatory gene expression in the cells of the presomitic mesoderm (PSM), a region at the tail end of the vertebrate embryo, thus generating regular patterns of somites (embryonic organs that develop into vertebrae and other mammalian repetitive structures (Gonzales and Kageyema, 2007)).

In mammals there are four known Notch genes that encode transmembrane receptors for mediating short-range signaling events. The five known ligands of Notch (Jagged-1,-2 and Delta like-1, -3, and -4) are also transmembrane proteins. At the cell surface, a Notch receptor can interact with one of its ligands in a neighboring cell leading to the release of the Notch intracellular domain (NICD). The subsequent nuclear translocation of NICD results in transcriptional activation of specific genes (Hes and Her/Hesr families) whose corresponding proteins in turn act as transcriptional repressors. There is evidence that endogeneous NICD acts at very low concentration (Fiúza and Arias, 2007), strongly suggesting a stochastic simulations approach for modeling Delta-Notch signaling. In both models, the transcriptional and translational delays are responsible for the oscillatory behavior. The involved genetic regulation is modeled by delayed Hill type reactions.

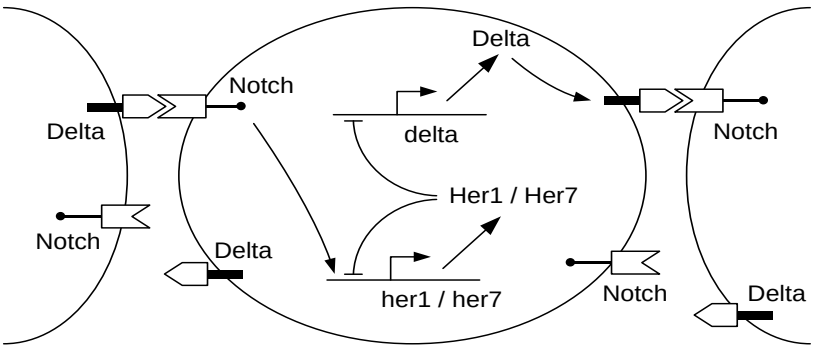


Figure 2: Delta-Notch signaling pathway and the autoinhibition of Notch target genes her1 and her7. Delta proteins in the neighboring cells activate the Notch signal within the cell.

4.1 Delta-Notch Signaling: Hes1 and Her1/7

4.1.1 Hes1

The hes1 gene is one of the best characterized genes in the segmentation clocks. Hirata et al. (2004) measured the production of hes1 mRNA (M) and Hes1 protein (P) in mouse. Serum treatments on cultured cells, that have already been shown to induce circadian

oscillation by Balsalobre et al. (1998), result in oscillations in expression levels for hes1 mRNA and Hes1 protein in a two hour cycle. Between the oscillatory profiles of mRNA and protein is a phase lag of approximately 15 min. The oscillations in expression continue for 6–12 h and are not dependent on the stimulus but can be induced by exposure to cells expressing Delta. It has been argued that the lag between protein and mRNA oscillation levels of 15 min reflects the time needed for protein degradation. Specifically, the data presented in the paper by Hirata et al. (Figure 1 in Hirata et al., 2004) indicates sustained oscillation of hes1 mRNA over six periods and that oscillation of Hes1 protein that dies away after 6–8 h.

Hirata et al. examined the underlying mechanisms for the observed oscillations and showed that in the presence of the proteasome inhibitor MG132, hes1 mRNA is initially induced but after 3 h it is suppressed because of constant repression of transcription by persistently high protein levels (negative autoregulation). Treatment with cycloheximide leads to sustained increase of hes1 mRNA and blocks its oscillation. A similar effect occurs with overexpression of dnHes1, a dominant-negative form of Hes1 that is known to suppress Hes1 protein activity (Ström et al., 1997). These results reveal that both Hes1 protein synthesis and degradation are needed for oscillations in the expression levels of hes1 mRNA. Other experiments showed that the same mechanisms hold for hes1 mRNA expression levels in the PSM in mouse. Hirata et al. also estimate the half-lives of hes1 mRNA and Hes1 protein to be 24.1 +/- 1.7 min, 22.3 +/- 3.1 min, respectively. Experiments with various protease inhibitors suggest that Hes1 protein is specifically degraded by the ubiquitin–proteasome pathway.

Since the simple negative feedback loop of hes1 mRNA and Hes1 was unable to generate sustained oscillations when modeled as a system of two ODEs, Hirata et al. postulated a Hes1 interacting factor as a third molecular species. Subsequently, they obtained a system of three ODEs that was then able to generate sustained oscillatory behavior. However, there is no direct experimental evidence for such an interacting factor.

Later, it was shown that simple coupled delay differential equations (DDEs), representing the time delays due to transcription and translation, are able to explain the sustained oscillations without recourse to the addition of a third variable (Monk, 2003; Jensen et al., 2003; Lewis, 2000; Bernard et al., 2006). Monk and Jensen et al. proposed the DDE

$$\begin{aligned} \frac{dM}{dt} &= \alpha_M f\left(P(t-\tau)\right) - \mu_M M \\ \frac{dP}{dt} &= \alpha_P M(t) - \mu_P P \end{aligned}$$

for the two species, hes1 mRNA (M) and Hes1 (P) and a regulatory Hill function

$$f(P(t)) = \frac{1}{1+(P(t-\tau)/P_0)^{\mathrm{h}}}$$

representing the repression of mRNA production by the binding of Hes1 dimers to the promoter region, with combined transcriptional and translational delay τ, Hill coefficient h and DNA dissociation constant P_0. The reaction rates μ_M and μ_P are the degradation rates of hes1 mRNA and Hes1, respectively, α_M is the maximal mRNA transcription rate in the absence of protein repression, and α_P is the translation rate. See Table II for parameters.

Jensen et al. showed via simulations that for the case $h = 2$, oscillations are only sustained for $\tau > 80$ and there are no oscillations for $\tau < 10$. For $\tau \in (10,80)$, the period of the damped oscillations is approximately 170 min, which is much greater than the observed period of 120 min. Bernard et al. had shown previously for a modification of the DDE model

by Monk that for the experimentally observed period of T=120 min, sustained oscillations can only be obtained for $h \geq 4.1$, $\tau \geq 19.7$. On the other hand, it was argued that since the transcription factor is a Hes1 dimer and there are at least three separate binding sites for Hes1 dimers in the regulatory region of the hes1 gene, an appropriate value of h is at least 2. However, whether h should be as large as 4.1 is debatable.

Barrio et al. (2006) studied the Hes1 negative feedback loop as a discrete, stochastic delay model based on the DDE model by Monk (2003). The chemical kinetics is described by the following reactions:

$$R_1: \quad M \xrightarrow{\mu_M} \emptyset,$$
$$R_2: \quad P \xrightarrow{\mu_P} \emptyset,$$
$$R_3: \quad M \xrightarrow{\alpha_P} M + P,$$
$$R_4: \quad P \xrightarrow{\alpha_M, f, \tau} M + P.$$

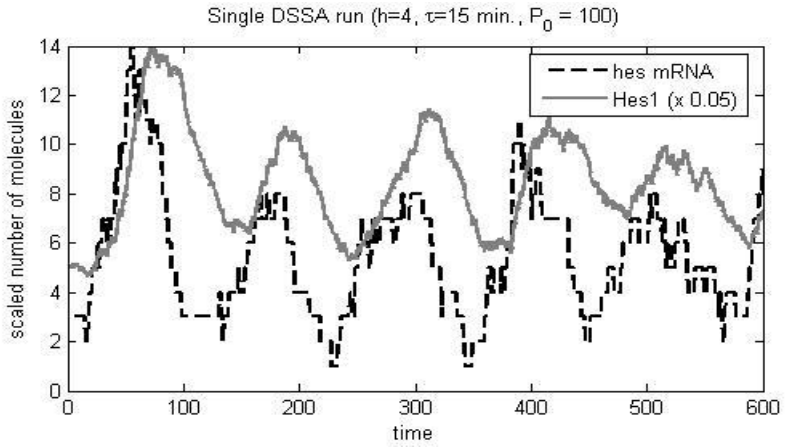
Reactions R_1 and R_2 are the degradations of M and P, respectively. R_3 represents the translation of M and R_4 is the regulated transcription with Hill function f.

By performing discrete stochastic simulations of the model with varying values for h, τ, and P_0 using the DSSA algorithm, Barrio et al. showed that h need not be as large as 4.1 to obtain sustained oscillations when discrete models are used. The results indicate that in the presence of intrinsic noise the critical value of the Hill coefficient, under which the system dynamics does not show sustained oscillations, decreases to just less than 3. Reasonably well-defined sustained regular oscillations could be observed for values of $\tau = 15$ with $h = 4$, and $\tau = 10$ with $h = 3$ (Figure 3). Values for τ lower than 10 result in noisy and irregular delay. By knowing more accurate values for the transcriptional and translational delays an even more accurate prediction of h might be possible and vice-versa.

parameter	description	value	Reference
μ_M	Hes1 mRNA degradation rate	0.029 [min^{-1}]	Hirata et al. (2002)
μ_P	Hes1 degradation rate	0.031 [min^{-1}]	Hirata et al. (2002)
α_P	translation rate	1 [min^{-1}]	Monk (2003)
α_M	max. transcription rate	1 [min^{-1}]	Normalized; Monk (2003)
P_0	critical no. of Hes1 protein (Hill function parameter)	10-100	Lewis (2003), Monk (2003)
h	Hill cooperativity factor (Hill function parameter)	2-4	Lewis (2003), Monk (2003)
τ	total delay (transcription, translation, translocation)	10-40 [min]	Monk (2003)

Table II: Parameters used in the Hes1-model.

(a)



(b)

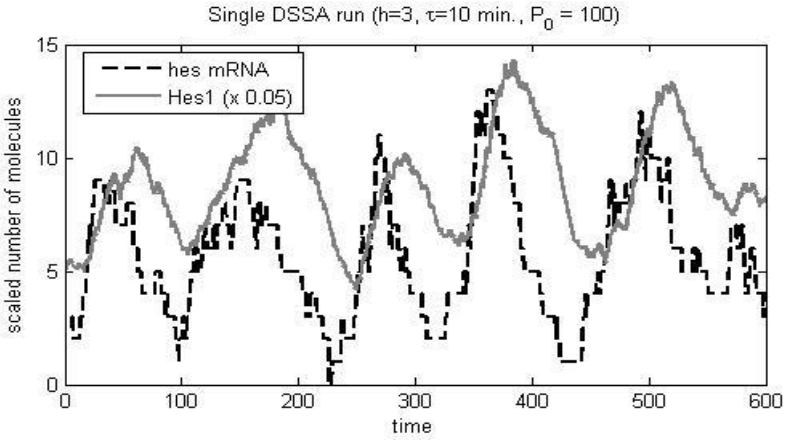


Figure 3: Single DSSA trajectories for values of (a) $\tau = 15\,\text{min}$ with $h = 4$, and (b) $\tau = 10\,\text{min}$ with $h = 3$ ($P_0 = 100$).

Barrio et al. (2006) computed the arithmetic mean over 1,000 independent stochastic simulation runs for constant and variable delay. In spite of the differences between individual simulations due to inherent stochasticity, the arithmetic mean showed damped oscillation. This matched the biological experiments where Western-blots of Hes1 from the whole cell population showed damped oscillations that are arrested after eight hours. However, the difference between individual stochastic simulations and the mean suggests that the damping, observed at the whole population level, arises from desynchronization of Hes1 oscillation in individual cells. This was supported by real-time imaging experiments showing that the oscillations in individual cells continue for longer than 8 hours (Masamizu et al., 2006).

The study of the Hes1 negative feedback loop demonstrated the usefulness of the DSSA for chemical kinetics involving delays. Because this approach is very general, it is able to provide deep insights into the relationship between delayed processes, intrinsic noise, and small numbers of molecules in many biological systems.

4.1.2 Her1/7

The Notch signaling pathway, which includes several signaling molecules (such as Hes1 and Her1/Her7) in mouse and zebrafish, respectively, plays a key role in the segmentation clock of vertebrates. In (wildlife) zebrafish, about 30–32 somites are formed at a rate of one every 30 min (at 28±C). Although it is suggested that some anterior somites (12) are derived

due to some form of dorsal convergence, most somites emerge sequentially from the PSM. It is distinguished between the posterior and anterior parts of the PSM. In zebrafish embryos at a developmental stage of 10 somites, the posterior PSM extends over 25 cells in anterior to posterior axis, which are the precursors for approximately five somites, each about five cells in length. The anterior PSM contains the cells that lead to the next two to three somites.

In zebrafish, the genes her1 and her7 are autorepressed by their own gene products (Her1 and Her7) and positively regulated by Notch signaling (Lewis 2003; Giudicelli and Lewis, 2004) - Figure 2. In both cases, transcriptional and translational delays are responsible for the oscillatory behavior and determine its period. Additional information on the somite segmentation clock in zebrafish is in Holley (2007) and Lewis and Ozbudak (2007).

Horikawa et al. (2006) performed experiments in which they investigated the system level properties of the segmentation clock in zebrafish. Their main conclusion is that the segmentation clock behaves as a coupled oscillator. The key element is the Notch-dependent intercellular communication, which is regulated by the internal hairy oscillator and whose coupling of neighboring cells synchronizes the oscillations. In one particular experiment, they replaced coupled cells by cells that were out of phase with the remaining cells and showed that at a later stage they still became fully synchronized. Clearly, the intercellular coupling plays a crucial role in minimizing the effects of noise to maintain coherent oscillations.

The stochastic model is based on the chemical reaction models by both Lewis (2003) and Horikawa et al. (2006). Lewis models a single cell and two coupled cells. His work is generalized by Horikawa et al. to a one-dimensional array of *n* cells. For each cell we simulate the dynamics of 6 different species controlled by 12 reactions. Denote by M_{h1_i}, M_{h7_i}, M_{d_i}, P_{h1_i}, P_{h7_i}, and P_{d_i} the species Her1 mRNA, Her7 mRNA, DeltaC mRNA, Her1 protein, Her7 protein and DeltaC protein in a particular cell *i*. For each of the species $S = M_{h1_i}, M_{h7_i}, M_{d_i}, P_{h1_i}, P_{h7_i}, P_{d_i}$, the model contains a degradation reaction

$$S \xrightarrow{c} \emptyset$$

with associated rate constant $c = c_{h1}, c_{h7}, c_d, b_{h1}, b_{h7}, b_d$. The three different proteins P_{h1_i}, P_{h7_i}, P_{d_i} are synthesized with translational delays τ_{h1p}, τ_{h7p}, and τ_{dp}, respectively. The corresponding reactions are

$$S_1 \xrightarrow{c} S_1 + S_2$$

with $(S_1, S_2) = (M_{h1_i}, P_{h1_i})$ or (M_{h7_i}, P_{h7_i}) or (M_{d_i}, P_{d_i}) and associated reaction rate constants $c = a_{h1}, a_{h7}, a_d$. The transcription of M_{h1_i}, M_{h7_i} and M_{d_i} are regulated reactions with transcriptional delays τ_{h1m}, τ_{h7m}, and τ_{dm}, respectively. The reactions are

$$S_1 \xrightarrow{cf} S_1 + S_2$$

with $(S_1, S_2) = (P_{h1_i}, M_{h1_i})$ or (P_{h7_i}, M_{h7_i}) and associated reaction rate constants $c = k_{h1}, k_{h7}$

$$S_1 \xrightarrow{cg} S_1 + S_2$$

with $(S_1, S_2) = (P_{d_i}, M_{d_i})$ and $c = k_d$. As described in detail in Horikawa et al. (2006), the individual negative and positive regulations are modeled using specific Hill functions f and g. For cells *i* with $1 < i < n$ (all except for the first and last in the one-dimensional cell array) the Hill function f is defined by

$$f(P_{h1_i}, P_{h7_i}, P_{d_{i-1}}, P_{d_{i+1}}) = r_h \frac{1}{1 + P_{h1_i} P_{h7_i} / P_0^2} + r_{hd} \frac{1}{1 + P_{h1_i} P_{h7_i} / P_0^2} \frac{P_{d_{i-1}} + P_{d_{i+1}}}{2D_0 + P_{d_{i-1}} + P_{d_{i+1}}},$$

and for cell 1 and *n* it is given by

$$f\left(P_{h1_1}, P_{h7_1}, P_{d_1}\right) = \frac{1}{1+P_{h1_i}P_{h7_i}/P_0^2}\frac{P_{d_1}/D_0}{1+P_{d_1}/D_0}$$

$$f\left(P_{h1_n}, P_{h7_n}\right) = \frac{1}{1+P_{h1_n}P_{h7_n}/P_0^2}\frac{1}{1+D_0/500},$$

respectively. The parameters r_h and r_{hd} are weight parameters that determine the balance of internal and external contribution of oscillating molecules. Here, we assume 100% coupling, i.e. $r_{hd}=1$. For all cells, the Hill function g that describes the inhibition of DeltaC mRNA synthesis by Her1 and Her7 is given by

$$g\left(P_{h1_i}, P_{h7_i}\right) = \frac{1}{1+P_{h1_i}P_{h7_i}/P_0^2}.$$

The single cell, single-gene model consists only of 2 species (her1 mRNA and Her1 protein) and 4 reactions. The two degradation and the single translation reactions correspond to those in the *n*-cell model. For the inhibitory regulation of transcription a Hill function with Hill coefficient 2 is assumed (P_{h1} acts as a dimer). The Hill function takes the form

$$f\left(P_{h1}\right) = \frac{1}{1+P_{h1}/P_0^2}.$$

See Table III for the full list of model parameters.

A comparison of the DDE solutions with stochastic simulation results of the DSSA and Bτ-DSSA in Leier et al. (2007) and Burrage et al. (2007) revealed differences in the system dynamics. For a single cell, after an initial overshoot, the DDE solution shows completely regular amplitudes and an oscillatory period of approximately 40 minutes (Figure 4). In the intrinsic noise case there are still sustained oscillations but there is some irregularity in the profiles and the oscillatory period is closer to 50 minutes. The time lag (5-7 min) between protein and mRNA is about the same in both cases (Figure 5).

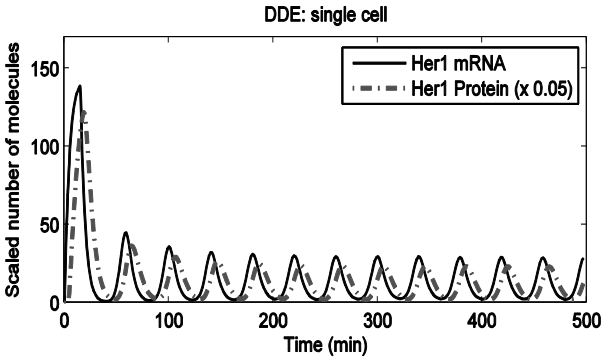


Figure 4: DDE solution for the Her1/Her7 single cell model

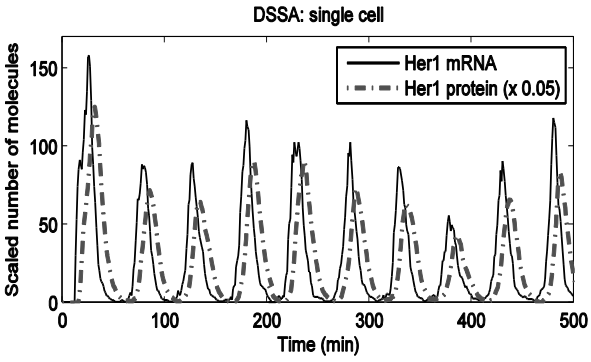


Figure 5: DSSA run for the Her1/Her7 single cell model

DSSA simulations of a one-dimensional array of 5 cells exhibit a period of oscillation that is closer to 45 minutes (Figure 6-7). The lag between protein and mRNA is about 25 minutes for DeltaC and about 7 minutes for Her1. Obviously, the cell coupling has some effect on the period of oscillation.

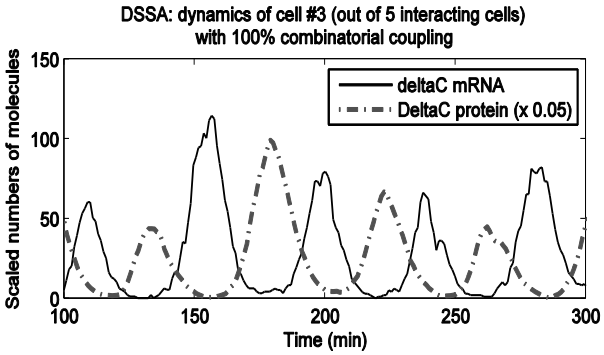


Figure 6: DSSA simulation of five Delta-Notch coupled cells , showing the dynamics of deltaC mRNA and protein in cell three

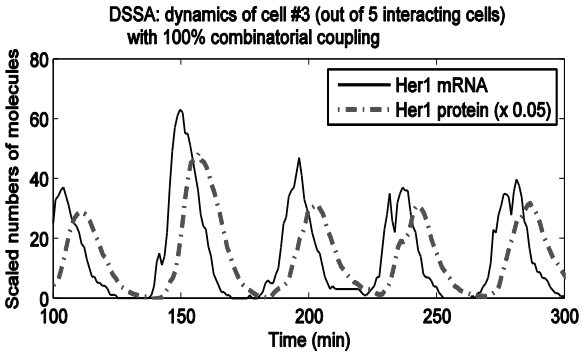
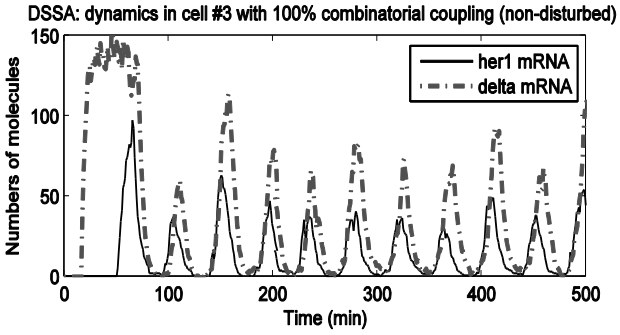
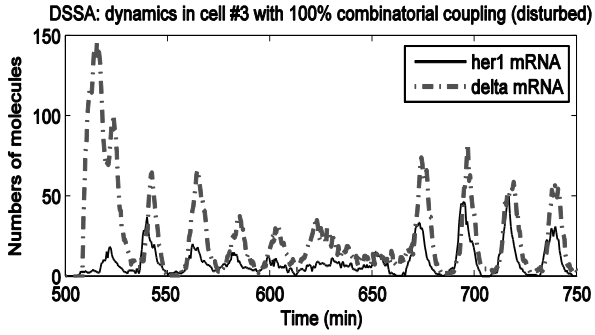


Figure 7: DSSA simulation of five Delta-Notch coupled cells , showing the dynamics of Her1 mRNA and protein in cell three

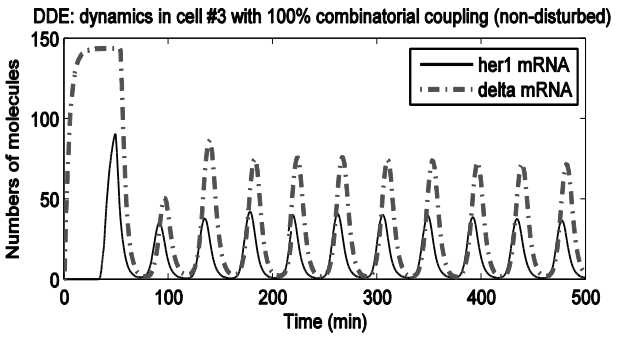
Leier et al. mimic an experiment by Horikawa et al. In both the DDE and the DSSA setting cell 3 (out of 5) is disturbed after a certain time period: after 500 minutes in the DSSA case and 260 minutes in the DDE case, at times when the delta mRNA levels are near their maximum. This is done by resetting all the values for cell 3 to zero at this point. This is meant to represent the experiment of Horikawa et al. in which some of the cells are replaced by oscillating cells that are out of phase. Horikawa et al. observed that nearly all the cells become resynchronized after three oscillations (90 min).



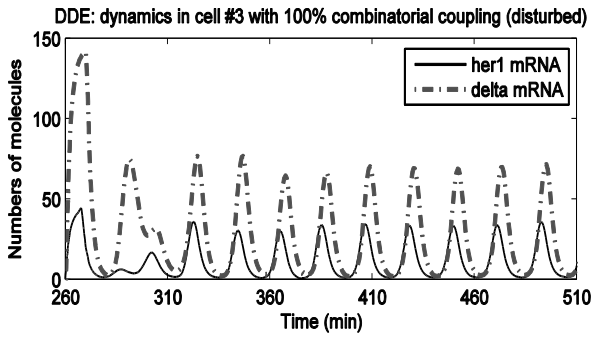
(a)



(b)



(c)



(d)

Figure 8: DSSA simulation result and DDE solution for the 5-cell array in the non-disturbed and disturbed setting. The graphs show the dynamics of deltaC and her1 mRNA in cell three. (a,c) DSSA and DDE results in the non-disturbed setting, respectively. (b,d) DSSA and DDE results in the disturbed setting. Initial conditions for cell 3 are set to zero. All other initial molecular numbers stem from the non-disturbed DSSA and DDE results in (a,c) after 500 and 260 minutes, respectively

In the DDE setting it takes about 60 minutes for the onset of resynchronization while in the DSSA setting it takes about 180 minutes (Figure 8). The difference can be partly due to the larger number of cells that are experimentally transplanted as well as differences in the cell arrangement between the three-dimensional *in vivo* experiments and the simulated one-dimensional cell array.

parameter	description	value
b_{h1}, b_{h7}, b_d	Her1/Her7/DeltaC protein degradation rate	0.23 [min^{-1}]
c_{h1}, c_{h7}, c_d	Her1/Her7/DeltaC mRNA degradation rate	0.23 [min^{-1}]
a_{h1}, a_{h7}, a_d	Her1/Her7/DeltaC protein synthesis rate (max.)	4.5 [min^{-1}]
k_{h1}, k_{h7}, k_d	Her1/Her7/DeltaC mRNA synthesis rate (max.)	33 [min^{-1}]
P_0	critical no. of Her1+Her7 protein/cell	40
D_0	critical no. of Delta protein/cell	1000
$\tau_{h1m}, \tau_{h7m}, \tau_{dm}$	time to produce a single Her1/Her7/DeltaC mRNA molecule	12.0, 7.1, 16.0 [min]
$\tau_{h1p}, \tau_{h7p}, \tau_{dp}$	time to produce a single Her1/Her7/ DeltaC protein	2.8, 1.7, 20.5 [min]

Table III: Parameters for the multicellular Her1-Her7 model. Parameter values are taken from Horikawa et al. (2006)

This study, although in an early stage, is another example indicating the relevance of both intrinsic noise delay models and continuous deterministic delay models for genetic regulatory systems. Despite some similarities between the dynamics of both the deterministic and stochastic models, the intrinsic noise simulations do make some predictions that are different from the deterministic model and that could be verified experimentally.

The reason for limiting the stochastic model to 5 cells is due to the long runtime of individual simulations when using the DSSA. To overcome the issue of small step-sizes, Leier et al. (2008(a)) introduced the Bτ-DSSA (see Section 3.4.2). The significant speed-up (while performing equally accurate as normal DSSA) allows the role of intrinsic noise and delay to be studied for large cellular systems and long time frames. There are many other issues that must be addressed when modeling both delays and intrinsic noise, one of which is how we represent delays. Clearly if delays are to represent complex processes such as transcription and translation, the delays should not be fixed but distributed. Appropriate distributions from which to sample the delays include uniform or truncated normal over some appropriate interval that represents lower and upper bounds for the delays. Other issues include whether it is appropriate to lump delays together into a single delay and how spatial effects associated with, for example, diffusion can be captured in purely temporal models by the use of delays.

5. Conclusions and Future Directions

In cell biology, cell signaling pathway problems are often tackled with a mix of deterministic temporal models, well mixed stochastic simulators, and/or hybrid methods. But, in fact, three dimensional stochastic spatial modeling of reactions happening inside the cell is sometimes needed in order to fully understand these cell signaling pathways. This is because

noise effects, low molecular concentrations, and spatial heterogeneity can all affect the cellular dynamics. However, there are ways in which important effects can be accounted without going to the extent of using these highly resolved spatial simulators. This reduces the overall computation time significantly, while at the same time still being able to capture the essential dynamics.

In this Chapter we have focused on how we can model both intrinsic noise and delayed reactions in a genetic regulatory setting via generalizations of the Stochastic Simulation Algorithm (the DSSA). We have also shown how we can coarsen in both time and space and demonstrated that this can improve the computational performance by several orders of magnitude over the DSSA. We have also shown, through two important applications, why we need algorithms that mimic both noise and delay effects as these approaches can capture the individual cell variability. We have also discussed what form the delays should take: fixed, variable, distributed, etc.

In the delay setting at least, codes based on the algorithms described here are still in their infancy and there is a need to standardize implementations and make these codes available to researchers. Future research must surely focus on multi-scale simulations and there is a great need to develop efficient algorithms that link different temporal and spatial scales – such as genetic regulatory models with those for cellular and organ function. This scientific field is wide open and can promise the dedicated researcher fascinating and rewarding endeavors.

References

an der Heiden U. Delays in physiological systems. J. Math. Biol. (1979), 8, 345-364.

Balsalobre A, Damiola F, Schibler U. A serum shock induces circadian gene expression in mammalian tissue culture cells. Cell (1998) 93, 929–937.

Barrio M, Burrage K, Leier A et al. Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation. PLoS Comp. Bio. (2006), 2(9), e117.

Burrage K, Burrage PM, Leier A et al. Stochastic delay models for molecular clocks and somite formation. Proceedings of SPIE (2008), 68020Z.

Burrage K, Hegland M, MacNamara S et al. A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems. Proceedings of the Markov 150th Anniversary Conference, 2006, Eds. AN Langville, WJ Stewart, 21-38, Boson Books.

Bernard S, Čajavec B, Pujo-Menjouet L et al. Modeling transcriptional feedback loops: The role of Gro/LTE1 in hes1 oscillations. Phil. Transact. A Math. Phys. (2006), Eng. Sci. 364, 1155-1170.

Bratsun D, Volfson D, Tsimring LS et al. Delay-induced stochastic oscillations in gene regulation. Proc. Natl. Acad. Sci. USA (2005), 102, 14593-14598.

Cai X. Exact Stochastic Simulation of Coupled Chemical Reactions with Delays. J. Chem. Phys. (2007), 126 (12), 124108-124116.

Cao Y, Gillespie DT, Petzold LR. Avoiding Negative Populations in Explicit Tau Leaping. J. Chem. Phys. (2005), 123, 054104.

Cao Y, Gillespie DT, Petzold LR. Efficient Stepsize Selection for the Tau-Leaping Method. J. Chem. Phys. (2006), 124, 044109.

Chatterjee A, Vlachos DG. Multiscale spatial Monte Carlo simulations: Multigriding, computational singular perturbation, and hierarchical stochastic closures. J. Chem. Phys. (2006), 124, 064110.

Chatterjee A, Vlachos DG. Temporal acceleration of spatially distributed kinetic Monte Carlo simulations. J. Comp. Phys. (2006), 211 (2), 596–615.

Chatterjee A, Vlachos DG, Katsoulakis MA. Binomial distribution based τ-leap accelerated stochastic simulation. J. Chem. Phys. (2005), 124, 044109.

Elf J, Ehrenberg M. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. Systems Biology (2004), 2, 230.

Elf J, Doncic A, Ehrenberg M. Mesoscopic reaction-diffusion in intracellular signaling. Proc. SPIE (2003), 5110, 114-124.

Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. Nature (2000), 403, 335–338.

Fiúza U-M, Arias AM. Cell and molecular biology of Notch. J. Endocrinology (2007), 194, 459–474.

Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. J. Phys. Chem. A (2000), 104, 1876-1889.

Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem. (1977), 81, 2340-2361.

Gillespie DT. The chemical Langevin equation. J. Chem. Phys. (2000), 113, 297-306.

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

Gillespie DT, Petzold LR. Improved leap-size selection for accelerated stochastic simulation. J. Chem. Phys. (2003), 119 (16), 8229-8234.

Giudicelli F, Lewis J. The vertebrate segmentation clock. Curr. Opin. Genet. Dev. (2004), 14(4), 407–414.

Gonzales A, Kageyema R. Practical Lessons from Theoretical Models about the Somitogenesis. Gene Regulation and Systems Biology (2007), 1, 35-42.

Goodwin BC. Oscillatory behavior in enzymatic control processes. Adv. Enzyme Regul. (1965), 3, 425-438.

Hattne J, Fange D, Elf J. Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics (2005), 21, 2923.

Hirata H, Yoshiura S, Ohtsuka T et al. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. Science (2002), 298, 840–843.

Holley SA. The genetics and embryology of zebrafish metamerism. Dev. Dyn. (2007), 236(6), 1422–1449.

Horikawa K, Ishimatsu K, Yoshimoto E et al. Noise-resistant and synchronized oscillation of the segmentation clock. Nature (2006), 441(7094), 719–723.

Jensen MH, Sneppen K, Tiana G. Sustained oscillations and time delays in gene expression of protein Hes1. FEBS Lett. (2003), 541, 176–177.

Kurtz TG. The relationship between stochastic and deterministic models of chemical reactions, J. Chem. Phys. (1972), 57, 2976-2978.

Leier A, Márquez-Lago T, Burrage K. Modeling intrinsic noise and delays in chemical kinetics of coupled autoregulated oscillating cells. Int. J. Multiscale Computational Engineering, (2008), 6(1).

Leier A, Márquez-Lago T, Burrage, K. Generalized binomial τ-leap method for biochemical kinetics incorporating both delay and intrinsic noise. J. Chem. Phys. (2008(a)), 128, 205107

Lewis J. Autoinhibition with transcriptional delay: A simple Mechanism for the Zebrafish somitogenesis oscillator. Curr. Biol. (2003), 13, 1398-1408.

Lewis J, Ozbudak EM. Deciphering the somite segmentation clock: beyond mutants and morphants. Dev. Dyn. (2007), 236(6), 1410–1415.

MacNamara S, Burrage K, Sidje R. Multiscale modeling of chemical kinetics via the master equation, SIAM J. Multiscale Modelling and Simulation Multiscale Modeling & Simulation (2008), Vol.6, No.4.

Marquez-Lago T, Burrage K. Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics. J. Chem. Phys. (2007), 127 (10), 104101.

Masamizu Y, Ohtsuka T, Takashima Y et al. Real-time imaging of the somite segmentation clock: Revelation of unstable oscillators in the individual presomitic mesoderm cells. PNAS (2006), 103 (5), 1313–1318.

Monk NAM. Oscillatory expression of Hes1, p53, and NF-κB driven by transcriptional time delays. Curr. Biol. (2003), 13, 1409-1413.

Morton-Firth CJ, Bray D. Predicting temporal fluctuations in an intracellular signalling pathway. J. Theor. Biol. (1998), 192, 117-128.

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

Nicolau Jr. DV, Burrage K, Parton RG et al. Identifying Optimal Lipid Raft Characteristics Required To Promote Nanoscale Protein-Protein Interactions on the Plasma Membrane. Mol. Cell. Biol. (2006), 26, 313-323.

Peng X, Zhou W, Wang Y. Efficient binomial leap method for simulating chemical kinetics. J. Chem. Phys. (2007), 126, 224109.

Rathinam M, Petzold LR, Cao Y, et al. Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method. J. Chem. Phys. (2003), 119, 12784.

Resat VH, Wiley HS, Dixon DA. Probability-weighted dynamic Monte Carlo method for reaction kinetics simulations. J. Phys. Chem. B (2001), 105, 11026-11034.

Ström A, Castella P, Rockwood J, et al. Mediation of NGF signaling by post-translational inhibition of HES-1, a basic helix–loop–helix repressor of neuronal differentiation. Genes Dev. (1997), 11, 3168–3181.

Tian T, Burrage K. Binomial leap methods for simulating stochastic chemical kinetics. J. Chem. Phys. (2004), 121 (21), 10356-10364.

Tian T, Burrage K, Burrage PM, Carletti M. Stochastic Delay Differential Equations for Genetic Regulatory Networks. J. Comp. Applied Maths. (2007), 205, 2, 696-707.

Turner T, Schnell S, Burrage K. Stochastic approaches for modelling in vivo reactions. Comput. Biol. Chem. (2004), 28, 165.

Additional Reading

Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. J. Chem. Phys. (2007), 127, 214107.

Anderson DF**.** Incorporating postleap checks in tau-leaping. J. Chem. Phys. (2008), 128, 054103.

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

Burrage K, Tian T, Burrage PM. A multi-scaled approach for simulating Chemical Reaction Systems. Prog. Biophys. Mol. Biol. (2004), 85, 217-234.

Bessho Y, Kageyama, R. Oscillations, clocks and segmentation. Curr. Opin. Genet. Dev. (2003), 13, 379–384.

El Samad H, Khammash M, Gillespie D. Stochastic Modeling of Gene Regulatory Networks. Int. J. Robust and Nonlinear Control. (2002), 15, 691-711.

Gillespie DT. Stochastic simulation of chemical kinetics. Ann. Rev. Phys. Chem. (2007), 58, 35-55.

Goutsias J. Classical versus stochastic kinetics modeling of biochemical reaction systems. Biophys. J. (2007), 92, 2350-2365.

Kitano H. Computational systems biology. Nature (2000), 420, 206-210.

Márquez Lago T, Leier A, Burrage K. Modeling molecular translocation processes with a stochastic delay simulation algorithm. In preparation for submission, 2008.

McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. Trends Genet. (1999), 15, 65-69.

Puchalka J, Kierzek AM. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. Biophys. J. (2004), 86, 1357-1372.

Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. Science (2005), 309, 2010-2013.

Saga Y, Takeda H. The making of the somite: molecular events in vertebrate segmentation. Nat. Rev. Genet. (2001), 2, 835-84.

Ullah M, Schmidt H, Cho K, et al. Deterministic modelling and stochastic simulation of biochemical pathways using MATLAB. Systems Biology, IEEE Proceedings (2006), 153, 53-60.

Zheng Q, Ross J. Comparison of deterministic and stochastic kinetics for nonlinear systems. J. Chem. Phys. (2001), 94, 3644-3648.