HAL
open science

Simulation of Gene Regulatory Networks

Bernard Ycart, Frédéric Pont, Jean-Jacques Fournié

To cite this version:

Bernard Ycart, Frédéric Pont, Jean-Jacques Fournié. Simulation of Gene Regulatory Networks. [Research Report] LJK. 2011. hal-00869111

HAL Id: hal-00869111

https://hal.science/hal-00869111

Submitted on 2 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Simulation of Gene Regulatory Networks

Bernard Ycart* Frédéric Pont† Jean-Jacques Fournié‡

Abstract

This limited review is intended as an introduction to the fast growing subject of mathematical modelling of cell metabolism and its biochemical pathways, and more precisely on pathways linked to apoptosis of cancerous cells. Some basic mathematical models of chemical kinetics, with emphasis on stochastic models, are presented.

Keywords: biochemical pathways; kinetics; master equation; diffusion processes; simulation
MSC: 80A30; 92C40

1 Introduction

This is intented as an introduction to the rapidly growing literature on the mathematical modelling of biochemical pathways. This subject links several, quite different areas of research. Depending on viewpoints, relevant articles can be found in Biology, Chemistry, Mathematics, or Informatics journals. Rather than seeking a hopeless exhaustivity, we have tried to illustrate some of the current approaches by a few recent references.

Inside the huge domain of biochemistry, we have focused on cell metabolism and its biochemical pathways, and more precisely on pathways linked to apoptosis of cancerous cells. On the other hand, mathematical modelling in genomics has used many different techniques, among which we chose to restrict our study to dynamic equations, insisting on stochastic models.

The paper is organized as follows. In the next section we shall review a few articles on biochemical pathways, enlighting the use of modelling and in silico experiments. Section 3 presents the basic mathematical models of chemical kinetics, with emphasis on stochastic models. Section 4 deals with the computer simulation of stochastic models, with emphasis on Gillespie's algorithm. Computer integration aspects are treated in section 5.

2 Biochemical pathways

In the ever growing literature on cell metabolism and biochemical pathways, we have selected a few references dealing with the targeting of regulatory pathways, in particular apoptosis pathways, and their mathematical modelling. The global human metabolic map has been reconstructed accounting for the functions of 1496 ORF's, 2004 proteins, 2766 metabolites, and 3311 metabolic and transport reactions [24]. A good review on apoptosis pathways is given by Elmore [26].

*Laboratoire Jean Kuntzmann, Univ. Grenoble-Alpes, France Bernard.Ycart@imag.fr
†INSERM UMR1037-Cancer Research Center of Toulouse, FranceFrederic.Pont@inserm.fr
‡INSERM UMR1037-Cancer Research Center of Toulouse, FranceJean-Jacques.Fournie@inserm.fr

Among the most recent advances, Folger *et al.* [30] announce a genome-scaled network model of cancer metabolism that they validate by predicting 52 cytostatic drug targets. Shaughnessy *et al.* [80] describe a way of artificially regulating MAP kinase cascades.

Clarke *et al.* [14] use statistical model checking for analysing t-cell receptor signalling pathways.

Studies linking biochemical pathways to cancer outcomes include that of Chuang *et al.* [13] and Taylor *et al.* [100].

Devun *et al.* [23] have recently identified a "mitochondrial permeability transition pore-dependent pathway to selective cell death". Teams who have used apoptosis pathways for cancer therapy include Chu and Chen [12], Ghobrial *et al.* [34], Speirs *et al.* [98]. A success obtained by combining Haem oxygenase with fumarate hydratase has recently been announced by Frezza *et al.* [31]. Yosef *et al.* [107] address the problem of optimizing gene networks, with application to apoptosis.

The ultimate goal to systematically predict through mathematical modelling, possible gene targets that could induce cancerous cell apoptosis has been pursued by several teams among which Legewie *et al.* [63], Ryu *et al.* [90], Huber *et al.* [51].

3 Mathematical models

Many mathematical models have been applied to genomics. The interested reader is referred to Shmulevitch and Dougherty's book [97] or the reviews by de Jong [16], Goutsias and Lee [44], Karlebach and Shamir [56]. Here we will focus exclusively onto the most classical dynamic models which are those of chemical kinetics. The mathematical developments date back to the first half of the twentieth century (see the historical section of [74]), and are presented in many textbooks, in particular those of van Kampen [103], Ethier and Kurtz [27], Gardiner [32], Wilkinson [104], or Kolokoltsov [60]. A short and clear introduction has been written by Higham [49], from which we shall borrow the terminology regarding the main three "chemical equations". As examples, we shall use the simplest historic models of Yule [108] and Michaelis-Menten [75, 54].

At the basis of the theory, modelling assumptions are made to ensure that the only quantities of interest be the number of molecules of each type of reactant simultaneously present: constant volume, well stirred medium, space homogeneity. As remarked by Higham [49] section 9.3, these hypotheses could be questioned when applied in the biological context of the living cell. Nevertheless, they are usually considered as unavoidable, and the models have been theoretically established on a very firm physical and chemical base: see for instance Gillespie [37].

A chemical reaction system is made of a certain number n of *reactants* (or chemical species), and another number m of *reactions* destroying or producing molecules of them. The n-dimensional *state vector* $X(t)$ keeps track of the quantities of each species simultaneously present at time t: its i-th entry counts how many molecules of species i are in the solution at time t. According to the scale of time and numbers, three different models are considered.

1. *Microscopic scale:* the vector $X(t)$ is a stochastic process with birth-and-death dynamics: reactions take place one at a time at random instants and modify $X(t)$ by a few units added or substracted from some coordinates. The probabilities of all possible states are the solution of the Chapmann-Kolmogorov equations for the Markov process

$\{X(t)\}$. That system of Ordinary Differential Equations (ODE's) is called the *Chemical Master Equation.*

2. *Mesoscopic scale:* when rescaling time and space, the Markov jump process $\{X(t)\}$ converges to a continuous real-valued diffusion process, solution of a system of Stochastic Differential Equations (SDE's), called the *Chemical Langevin Equation.*

3. *Macroscopic scale:* when stochastic fluctuations are neglected, the quantities of molecules are viewed as derivable functions of time, solution of a system of ODE's, called the *Reaction Rate Equations.*

The Chemical Master Equation can be explicitly solved only in exceedingly simple, mostly irrelevant cases, such as the Yule model that we shall treat below (other examples are reviewed by McQuarrie in section III of [74]). The alternative approach is to simulate trajectories of $X(t)$ through Gillespie's algorithm, that we shall examine in more detail in the next section. As pointed out in section 6 of Higham [49], the so called "tau-leaping" method, which is the usual way of accelerating the algorithm, can be viewed as an approximate solution to the Chemical Langevin Equation. On the theoretical side, convergence of birth-and-death dynamics to diffusion processes has been proved by Kurtz at the end of the 70's: chapter 10 of [27] gives a complete theoretical treatment of diffusion processes in chemical kinetics. Thus, from the modelling point of view, it can be considered that the microscopic and mesoscopic scales are not essentially different: discrepancies come from the mathematical or algorithmic treatment. The real opposition is between stochastic (microscopic and mesoscopic scales) and deterministic (macroscopic scale) modelling. It has been discussed at length in many references, see for instance Chen *et al.* [11], Goutsias [43], Liang and Qian [66], Qian [81] or Shahrezaei and Swain [95]. The main reason why stochastic is usually preferred to deterministic for gene regulatory networks lies in the order of magnitude of the numbers of molecules involved. The proteins produced by mRNA can usually be counted by a few hundreds per second, far short from Avogadro's number ($6.02\ 10^{23}$) which is the scale at which deterministic models work.

We shall now introduce the main concepts of mathematical kinetics models, illustrating them on the two basic examples of the Yule and Michaelis-Menten models. The effect of reactions on species is usually described by symbolic equations of the type $A + B \longrightarrow C$, meaning that each time the reaction takes place or *fires*, one molecule of A and one molecule of B combine to form one molecule of C. For mathematical purposes, it is convenient to encode the effect of each reaction by two vectors of integers, called the *stoichiometric vectors* (from the Greek meaning "measure of elements"). To the j-th reaction correspond vectors ν_j^- and ν_j^+. Their entries are as follows.

- $\nu_j^-(i)$ is the number of molecules of species i that are *destroyed* by reaction j,
- $\nu_j^+(i)$ is the number of molecules of species i that are *produced* by reaction j.

For instance reaction $A + B \longrightarrow C$ will be translated by two 3-dimensional vectors, with entries indexed by A B and C.

$$\nu^- = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \nu^+ = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

The *Yule model* is a very basic population model used for instance in the case of bacteria population growth. Formally, it could be assimilated to a chemical equation with only one reactant: $A \longrightarrow 2A$ (one bacteria becomes two when the reaction/meiosis fires). The stoichiometric vectors only have one entry: $\nu^- = 1$ and $\nu^+ = 2$.

The *Michaelis-Menten* equations involve four species:

- a substrate S
- an enzyme E
- a complex C
- a product P

Here are the three equations and the corresponding stoichiometric vectors, indexed by S, E, C, P in that order.

1. $S + E \longrightarrow C$: $\quad \nu_1^- = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \nu_1^+ = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$

2. $C \longrightarrow S + E$, $\quad \nu_2^- = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \nu_2^+ = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$

3. $C \longrightarrow E + P$, $\quad \nu_3^- = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \nu_3^+ = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$

Some textbooks use the more compact *stoichiometric matrix* that summarizes in its columns the vectors above. Here is that matrix for the Michaelis-Menten model.

1	2	3	
−1	+1	0	S
−1	+1	+1	E
+1	−1	−1	C
0	0	+1	P

There are two reasons to prefer the vectors ν^+ and ν^-. One is that the matrix looses information when the same species appears on both sides of the equation (as in the Yule model). The other is that the propensities, to be defined later, depend only on the ν_j^-'s and not on the ν_j^+'s.

Regarding the mathematical expression of the different models, we shall follow the introduction of Ball *et al.* [3]. For the stochastic version of Michaelis-Menten dynamics, see Qian [82], Sanft *et al.* [91] and Higham [49]. Once the stoichiometry of the system is known, the

evolution of the state vector only depends on successive reaction firings. Denote by $R_j(t)$ the number of times reaction j fires between 0 and t. Then:

$$X(t) = X(0) + \sum_{j=1}^{m} R_j(t)(\nu_j^+ - \nu_j^-) \ .$$

The $\{R_j(t)\}$'s are counting processes, whose instantaneous rates are called the *propensities*. Intuitively, during an interval of time $[t, t+\delta t]$ short enough to ensure that only one reaction will fire in that interval, the probability that reaction j fires should be proportional to the duration δt. The proportionality coefficient is the instantaneous rate of $R_j(t)$. Assuming the system is well stirred, all molecules are equally likely to be at any location at any time. So the probability that all molecules required for reaction j meet, should be proportional to the number of ways of finding these molecules together. Let x be the vector of integers giving the number of molecules of each species. The propensity of reaction j when the state vector is x is:

$$\lambda_j(x) = \kappa_j \frac{N}{N^{\sum_i \nu_j^-(i)}} \prod_i \nu_j^-(i)! \binom{x}{\nu_j^-(1)\dots\nu_j^-(n)} \ ,$$

where N is a scaling parameter, usually taken to be the volume of the system multiplied by Avogadro's number, and κ_j is a constant, depending only on the reaction (including of course temperature conditions). Rather than detailing the general formula above, we shall make its meaning clear on simple examples, involving only 2 species A and B, for which the number of molecules are denoted by a and b.

- $A \longrightarrow \cdots : \lambda_j(x) = \kappa_j\, a$
- $2A \longrightarrow \cdots : \lambda_j(x) = \kappa_j \frac{1}{N}\, a(a-1)$
- $A + B \longrightarrow \cdots : \lambda_j(x) = \kappa_j \frac{1}{N}\, ab$
- $2A + 3B \longrightarrow \cdots : \lambda_j(x) = \kappa_j \frac{1}{N^4}\, a(a-1)\, b(b-1)(b-2)$

Using propensities, the counting processes $\{R_j(t)\}$ can be written as:

$$R_j(t) = Y_j\left(\int_0^t \lambda_j(X(s))\,\mathrm{d}s\right) \ ,$$

where the Y_j are independent unit Poisson processes. Hence the *Stochastic Integral Equation* (SIE) defining X:

$$X(t) = X(0) + \sum_{j=1}^{m} Y_j\left(\int_0^t \lambda_j(X(s))\,\mathrm{d}s\right)(\nu_j^+ - \nu_j^-) \ . \tag{SIE}$$

The *Chemical Master Equation* (CME) is the forward Chapmann-Kolmogorov equation corresponding to the Markov process $\{X(t)\}$. It is a linear system of ODE's where the unknowns are the probabilities for $X(t)$ to be in each possible state and the propensities are the coefficients.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}[X(t)=x] \quad = \quad & -\left(\sum_j \lambda_j(x)\right)\mathbb{P}[X(t)=x] \\ & + \sum_j \lambda_j(x-\nu_j^+ + \nu_j^-)\,\mathbb{P}[X(t) = x - \nu_j^+ + \nu_j^-] \ . \end{aligned} \tag{CME}$$

Here is the CME for the Yule model $A \longrightarrow 2A$, denoting by $p_n(t)$ the probability that n molecules of A are present at time t.

$$\frac{\mathrm{d}p_n(t)}{\mathrm{d}t} = -\kappa n p_n(t) + \kappa(n-1)p_{n-1}(t) \ .$$

It is a well know fact that this equation admits an explicit solution: given an initial number of molecules of n_0, the number of molecules at time t follows the negative binomial distribution with parameters n_0 and $\mathrm{e}^{-\kappa t}$:

$$p_n(t) = \binom{n-1}{n_0-1} \mathrm{e}^{-\kappa n_0 t}(1-\mathrm{e}^{-\kappa t})^{n-n_0} \ .$$

The expectation of that distribution is $n_0\mathrm{e}^{\kappa t}$, and its variance is $n_0(1-\mathrm{e}^{-\kappa t})\mathrm{e}^{2\kappa t}$. As already mentioned, cases like this one are very rare and no explicit solution to the CME exists in general.

We shall now introduce the renormalization that leads to the Chemical Langevin Equation. Let x be a value of the state vector (counting molecules of each species). If N is the volume multiplied by Avogadro's number, then $c = \frac{1}{N}x$ is the vector of *concentrations* (expressed in moles per unit volume). If the coordinates of x are large, then the propensities can be expressed asymptotically as follows:

$$\lambda_j(x) \simeq N\kappa_j \prod_i c(i)^{\nu_j^-(i)} \ .$$

Hence we can define "macroscopic" propensities as:

$$\widetilde{\lambda}_j(c) = \kappa_j \prod_i c(i)^{\nu_j^-(i)} \ .$$

Recall the SIE defining X, and divide both members by N:

$$\frac{X(t)}{N} = \frac{X(0)}{N} + \sum_{j=1}^m \frac{1}{N} Y_j\left(\int_0^t \lambda_j(X(s))\,\mathrm{d}s\right)(\nu_j^+ - \nu_j^-) \ .$$

Replacing X by the concentration vector $C = \frac{1}{N}X$ gives:

$$C(t) = C(0) + \sum_{j=1}^m \frac{1}{N} Y_j\left(\int_0^t N\widetilde{\lambda}_j(C(s))\,\mathrm{d}s\right)(\nu_j^+ - \nu_j^-) \ .$$

Now the central limit theorem describes the asymptotics for the unit Poisson process (random counting with an average count of 1 per time unit). At (large) time Nu the process reaches Nu on average with fluctuations of order $\sqrt{N}$ described by a standard Brownian motion.

$$\lim_{n\to+\infty} \frac{Y(Nu) - Nu}{\sqrt{N}} = W(u) \ ,$$

where Y is a unit Poisson process, W is the standard Brownian motion and the limit is understood in distribution. So for each reaction j, we can replace $Y_j(Nu)$ by $Nu + \sqrt{N}W(u)$ and get the approximation:

$$\frac{1}{N} Y_j\left(\int_0^t N\widetilde{\lambda}_j(C(s))\,\mathrm{d}s\right) \simeq \int_0^t \widetilde{\lambda}_j(C(s))\,\mathrm{d}s + \frac{1}{\sqrt{N}} W_j\left(\int_0^t \widetilde{\lambda}_j(C(s))\,\mathrm{d}s\right) \ .$$

The definition of $C(t)$ at mesoscopic scale will thus be:

$$\begin{aligned} C(t) &= C(0) + \left(\int_0^t \sum_j \widetilde{\lambda}_j(C(s))(\nu_j^+ - \nu_j^-)\,\mathrm{d}s \right) \\ &\quad + \frac{1}{\sqrt{N}} \left(\sum_j W_j \left(\int_0^t \widetilde{\lambda}_j(C(s))\,\mathrm{d}s \right) (\nu_j^+ - \nu_j^-) \right) . \end{aligned}$$

This diffusion process is a solution to the following SDE, called *Chemical Langevin Equation* (CLE).

$$\begin{aligned} \mathrm{d}C(t) &= \left(\sum_j \widetilde{\lambda}_j(C(s))(\nu_j^+ - \nu_j^-) \right) \mathrm{d}t \\ &\quad + \frac{1}{\sqrt{N}} \left(\sum_j \sqrt{\widetilde{\lambda}_j(C(s))}(\nu_j^+ - \nu_j^-)\,\mathrm{d}W_j \right) . \end{aligned} \tag{CLE}$$

By neglecting the stochastic term in the CLE, we get the classical *Reaction Rate Equation* (RRE).

$$\frac{\mathrm{d}}{\mathrm{d}t} C(t) = \sum_j \widetilde{\lambda}_j(C(s))(\nu_j^+ - \nu_j^-) . \tag{RRE}$$

When the first member of the RRE vanishes, there remains a system of algebraic equations, relatively easy to solve, at least numerically. Its solution (which is a solution to the RRE *constant in time*) is called a *steady state*, or chemical equilibrium. Not all systems admit a steady state, but general conditions of existence have been found: see Feinberg [28]. Stiefenhofer [99], Thomson and Gudawardena [101, 102] study a theoretical approach to the system of algebraic equations involved in the computation of steady states in the context of biochemical systems.

As an illustration, we give below the CLE of our two examples. For the Yule model:

$$\mathrm{d}a(t) = \kappa a(t)\,\mathrm{d}t + \sqrt{\kappa a(t)}\,\mathrm{d}W$$

Notice that in this case, the RRE $\mathrm{d}a(t) = \kappa a(t)\mathrm{d}t$ gives $a(t) = a(0)\mathrm{e}^{\kappa t}$, in accordance with the expectation of $X(t)$ given above. This is a particular case: the expectation of the distribution solving the CME or the CLE is not the solution of the RRE in general. The relations between different types of models has been analyzed by Gillespie *et al.* [41].

Here are the macroscopic propensities for the Michaelis-Menten model.

1. $S + E \longrightarrow C$: $\widetilde{\lambda}_1 = \kappa_1 s(t)e(t)$,
2. $C \longrightarrow S + E$: $\widetilde{\lambda}_2 = \kappa_2 c(t)$,
3. $C \longrightarrow E + P$: $\widetilde{\lambda}_3 = \kappa_3 c(t)$.

Here is the CLE:

$$
\begin{aligned}
\mathrm{d}s(t) &= \Big(-\kappa_1 s(t)e(t) + \kappa_2 c(t)\Big)\,\mathrm{d}t \\
&\qquad + \frac{1}{\sqrt{N}}\Big(-\sqrt{\kappa_1 s(t)e(t)}\,\mathrm{d}W_1 + \sqrt{\kappa_2 c(t)}\,\mathrm{d}W_2\Big) \\
\mathrm{d}e(t) &= \Big(-\kappa_1 s(t)e(t) + \kappa_2 c(t) + \kappa_3 c(t)\Big)\,\mathrm{d}t \\
&\qquad + \frac{1}{\sqrt{N}}\Big(-\sqrt{\kappa_1 s(t)e(t)}\,\mathrm{d}W_1 + \sqrt{\kappa_2 c(t)}\,\mathrm{d}W_2 + \sqrt{\kappa_3 c(t)}\,\mathrm{d}W_3\Big) \\
\mathrm{d}c(t) &= \Big(+\kappa_1 s(t)e(t) - \kappa_2 c(t) - \kappa_3 c(t)\Big)\,\mathrm{d}t \\
&\qquad + \frac{1}{\sqrt{N}}\Big(\sqrt{\kappa_1 s(t)e(t)}\,\mathrm{d}W_1 - \sqrt{\kappa_2 c(t)}\,\mathrm{d}W_2 - \sqrt{\kappa_3 c(t)}\,\mathrm{d}W_3\Big) \\
\mathrm{d}p(t) &= \Big(\kappa_3 c(t)\Big)\,\mathrm{d}t \\
&\qquad + \frac{1}{\sqrt{N}}\Big(\sqrt{\kappa_3 c(t)}\,\mathrm{d}W_3\Big)\,.
\end{aligned}
$$

Even knowing that mathematical models have been established long ago on very sound theoretical grounds, one must remain aware of two major issues. The first one is the combinatorics of reaction systems involved in gene regulatory networks. The complete cell metabolism involves reactants and reactions by the tens of thousands. No computer program can solve so large systems of equations, deterministic or stochastic. In order to find shortcuts that limit the amount of calculations required, many different approaches have been tempted. Ball *et al.* [3] propose to take into account the very different scales of times at which reactions fire, and distinguish the slower components of a system from the faster. In the same vein, Goutsias [42] eliminates the effect of the faster reactions by using steady states. Grognard *et al.* [45] replace the essentially non linear models by piece-wise linear approximations, for which explicit solutions can be computed. Using the same technique, Ropers *et al.* [89] announce interesting stability results. Among many others, Mallavarapu *et al.* [69] use modularity to divide a full scale model into more tractable components. This approach consists of considering one module of interest, while assuming that the rest of the system has reached a steady state. Gunawardena [46] gives a graph-theoretical basis to modularity. De Jong and Page [19] consider steady states in the context of piece-wise linear approximations. Rao and Arkin [88] and McNamara *et al.* [73] couple the simulation of the CME with steady states approximations. Ramaswamy *et al.* [86] study the stochastic fluctuations around the steady state.

The other important issue is the estimation of parameters. The models presented in the previous section are based on the propensity functions, parametrized by the reaction contants κ_j. Estimating these constants is not an easy task. In the slighty different context of viral dynamics, Miao *et al.* [77] have recently reviewed the different approaches to that problem, called *identifiability*. The most obvious technique, already used by Michaelis and Menten (see [54]), is regression analysis on experimental data: see Jaqaman and Danuser [53] for a general review. Batt *et al.* [4] or de Jong and Ropers [20] propose alternative approaches to parametric estimation. Even in cases where only raw estimates of the reaction constants were known, some encouraging results show that approximated simulations can still account

at least qualitatively for experimental results: see Gutenkunst *et al.* [47] or Ropers *et al.* [89].

4 Stochastic Simulation Algorithms

As already pointed out, there exists a structural reason to prefer stochastic models to deterministic ones for gene regulatory network modelling. For that reason, and also because numerical solvers for ODE's are well know and implemented in all mathematical softwares, we shall not present them here. Solutions to the stochastic models (either the CME or the CLE) can only be obtained through Monte-Carlo methods, *i.e.* by simulating the stochastic processes under consideration. The basic method of simulation is *Gillespie's algorithm*, also called Stochastic Simulation Algorithm in some references. Proposed in the context of chemical reactions by Gillespie [35, 36], it is a particular case of the general method for simulating Markov Jump Processes in continuous time through their imbedded Markov chain. The basic version simulates a trajectory of the state vector $X(t)$ by translating the SIE introduced in the previous section:

$$X(t) = X(0) + \sum_{j=1}^{m} Y_j \left(\int_0^t \lambda_j(X(s))\,\mathrm{d}s \right) (\nu_j^+ - \nu_j^-)\ . \tag{SIE}$$

It should be understood as follows. At any time s, all m reactions have independent firing times, that of reaction j being exponentially distributed with parameter $\lambda_j(X(s))$. The next reaction will fire after a time which is the minimum of these random variables. From elementary properties of exponential random variables, it can be deduced that:

1. the next firing will occur after a random time, exponentially distributed with parameter $\lambda_\bullet(X(s)) = \sum_{j=1}^{m} \lambda_j(X(s))$.

2. it will be the firing of reaction j with probability $\dfrac{\lambda_j(X(s))}{\lambda_\bullet(X(s))}$.

The two elementary steps above are easily simulated. Any programming language has a Random function. Successive calls of that function output sequences of pseudo-random numbers that can be considered as realizations of independent random variables, uniformly distributed on the interval $[0, 1]$. Using Random, the two elementary steps of the simulation are easily programmed.

1. Knowing λ, to simulate a random variable exponentially distributed with parameter λ, output $-\log(\mathsf{Random})/\lambda$.

2. Knowing $p_1, \ldots, p_m$, to simulate a random index j with probability p_j,

   (a) compute cumulated probabilities $q_j = p_1 + \cdots + p_j$

   (b) compare Random to the q_j's

   (c) if Random is in the interval $[q_{j-1}, q_j]$, output j.

When reaction j fires, the state vector $X(s)$ is incremented using the stoichiometric vectors ν_j^- and ν_j^+. The pseudo-code of the Gillespie Algorithm can be written as follows.

Initialize
    Time scale $T \longleftarrow 0$
    State vector $X \longleftarrow X(0)$
Repeat
    Compute propensities $\lambda_j(X)$ for $j = 1, \ldots, m$
    Sum them to get $\lambda_\bullet(X)$
    Compute random delay: $S \longleftarrow -\log(\text{Random})/\lambda_\bullet(X)$
    Compute probabilities: $p_j = \dfrac{\lambda_j(X(s))}{\lambda_\bullet(X(s))}$
    Among all reactions, choose reaction j with probability p_j
    Update
        Time scale: $T \longleftarrow T + S$
        State vector: $X \longleftarrow X + \nu_j^+ - \nu_j^-$
Until end of simulation

Higham [49] provides a Matlab implementation of that algorithm. However, several reasons make it quite inefficient.

- Billions of steps should be simulated to get a trajectory comparable to experimental results. Even though each one of these steps is relatively cheap, the computer time required for a full scale simulation is prohibitive.

- Even with an arithmetic coprocessor, the log function required to compute the random delay between firings is more expensive than additions and multiplication. Using it at each step is time consuming, and essentially useless: from the central limit theorem, it is known that a sum of independent random variables is asymptotically distributed as a normal variable with same expectation and variance. The expectation of an exponential with parameter λ is $\frac{1}{\lambda}$ and its variance is $\frac{1}{\lambda^2}$. It is much more efficient to cumulate at each step $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$, then update the time scale only at regular intervals (say after 10^4 firings), with a normal variable.

- If the number of reactions is large (which is the case of gene regulatory networks), choosing a reaction with a given probability can be expensive. A lot of computer time is lost in updating probabilities and cumulating them. Moreover, comparing Random to the cumulated probabilities results in many useless tests, especially when the p_i's have very different orders of magnitude. A lot of computer time can be saved by ranking the $p_i's$ in decreasing order. But this can be efficient only when the same probability vector is used many times.

- At each step, the state vector is updated by a random vector which is equal to $\nu_j^+ - \nu_j^-$ with probability p_j. But again the central limit theorem can be used. After a large number of steps, the state vector has been updated by a sum of independent random variables, the expectation and variance of which can be cumulated to simulate a normal approximation only at the end of the loop.

- Updating the state vector and the propensities at each step is also time consuming and relatively inefficient, because when the numbers of molecules are large, changing them by a few units does not essentially modify the results.

The remarks above where made long ago and lead to the so called *tau-leaping* method (discussed by Gillespie himself in [38]) that consists of considering propensities as constant for a certain (relatively large) interval of time τ, and updating the state vector by a normal random variable only after that time τ. In the pseudo-code below, we fixed a number of firings L instead of fixing a time interval τ, but the difference is not essential. We shall not detail the simulation code for normal distributions, which is routine. We call $\mathsf{Normal}(E, V)$ a function simulating normally distributed random variables with vector expectation E and covariance matrix V.

Initialize
    Time scale $T \longleftarrow 0$
    State vector $X \longleftarrow X(0)$
    Iterations number L
Repeat
    Compute propensities $\lambda_j(X)$ for $j = 1, \ldots, m$
    Sum them to get $\lambda_\bullet(X)$
    Compute expectation and variance of time increment
        $E_{time} \longleftarrow E_{time} + L/\lambda_\bullet(X)$
        $V_{time} \longleftarrow V_{time} + L/\lambda_\bullet(X)^2$
    Compute probabilities: $p_j = \dfrac{\lambda_j(X(s))}{\lambda_\bullet(X(s))}$
    Compute expectation and covariance of state vector increments
        $E_{state} \longleftarrow \sum_j p_j(\nu_j^+ - \nu_j^-)$
        $V_{state} \longleftarrow \sum_j p_j(\nu_j^+ - \nu_j^- - E_{state})(\nu_j^+ - \nu_j^- - E_{state})^t$
        Multiply by loop length
            $E_{state} \longleftarrow L \times E_{state}$
            $V_{state} \longleftarrow L \times V_{state}$
    Update time scale: $T \longleftarrow T + \mathsf{Normal}(E_{time}, V_{time})$
    Update state vector: $X \longleftarrow X + \mathsf{Normal}(E_{state}, V_{state})$
Until end of simulation

As pointed out by Higham [49] the tau-leaping method matches the Euler-Maruyama scheme for the Chemical Langevin Equation:

$$\begin{aligned} \mathrm{d}C(t) &= \left(\sum_j \widetilde{\lambda}_j(C(s))(\nu_j^+ - \nu_j^-) \right) \mathrm{d}t \\ &\quad + \frac{1}{\sqrt{N}} \left(\sum_j \sqrt{\widetilde{\lambda}_j(C(s))}(\nu_j^+ - \nu_j^-)\, \mathrm{d}W_j \right) . \end{aligned} \tag{CLE}$$

It is a well know fact that the Euler-Maruyama scheme, like the Euler scheme for ODE's, tends to be numerically unstable. At the expense of a slight increase in computer time, one gets a better precision by using instead the Heun scheme. The book by Kloeden and Platen [59] is the indispensable reference on numerical schemes for SDE's.

Improvements on the Stochastic Simulation algorithms have generated many works, and Gillespie and his co-authors have been very active: Liu *et al.* [65] give a review of existing methods; Cao *et al.* [8] investigate numerical stability of leaping methods; Gillespie *et al.* [39, 40] address the problem of accelerating the simulation when propensities have very different orders of magnitudes. In [41, 105], the approximations to the Michaelis-Menten equations

are discussed, and the gain in computer time of leaping is evaluated. The same problem of multiple time scales has also been adressed by Ball *et al.* [3], Haseltine and Rawlings [48], Liu and Vanden-Eijden [67], McColluma *et al.* [72]. Other improvements include Bruck and Gibson's [7] study on simulation of large systems, and the works by Ramaswamy and his co-authors [83, 84, 85] on partial propensity. Elf and Ehrenberger [25] replace simulation by an evaluation of fluctuations around the solution of the RRE using linear noise approximations. Zhou *et al.* [110] have applied a stochastic algorithm to coupled reactions with delays. Recently, Zeron and Santillan [109] published a numerical study of a gene network with negative feedback regulation, including the stability of steady state.

5 Computer integration

The "Virtual Cell" of Gunawardena *et al.* in Harvard or de Jong [17] in Grenoble, is still a beautiful dream; yet significant steps toward its realization have been made.

Regarding biochemical pathways, the process of knowledge accumulation and sharing has long been very lively. The "Pathway Resource List", or "PathGuide" [1] currently lists more than 400 databases on biological pathways and molecular interaction, most of them freely accessible. Examples include BiGG [93] MetaCyc and BioCyc [10], KEGG [55], TRANSPATH [61], BioCyc [62], Reactome [71], PANTHER [76], PID [92]. Several databases now come with integrated environments that offer different vizualisations, like Cytoscape [96], or even simulate the archived models, like BioModels [64] or Expasy [33]. Kitano [57] or Huang *et al.* [50] review computational tools available in Systems Biology.

In the accumulation process, the standardization issue was raised very early. In 2000, Bader and Hogue [2] had already proposed BIND, a data specification adapted to pathways. In 2005, Cary *et al.* [9] listed 170 existing databases and called for standard exchange formats to successfully integrate data on a large scale. These were already under development, and they integrate standard programming tools for the simulation and treatment of gene networks and pathways: BioPAX [22] SBML [52], CellML [78], Biolingua [70]. Other programming toolboxes include COBRA [5], GNA [18], PATIKA [21], and DIZZY [87], the latest being more focused on stochastic simulation.

Together with software realizations, informatics theoretical researches attempted a formalization of languages and investigation procedures. Fisher and Henzinger [29] call "Executable Biology" that research area constructing computational models of biological systems. Kitano [58] proposes a graphical notation for biochemical networks. Danos and Laneve [15] invent an abstract language for formal proteins. Berthoumieux *et al.* [6] study the identification of network models though incomplete high throughput datasets. For testing and experimenting purposes, Lok and Brent [68] construct an automatic generation tool for cellular reaction networks. Monteiro *et al.* [79] propose a service oriented architecture for integrated treatment of networks. Schultz et al. [94] discuss the clustering of computational models based on semantic annotations. Yamamoto *et al.* [106] have developped an Artificial Intelligence system, SOLAR, that includes reasonning tools for biological inference on pathways.

References

[1] G.D. Bader, M.P. Cary, and C. Sander. Pathguide: a pathway resource list. *Nucleic Acid Res.*, 34:505–505, 2006.

[2] G.D. Bader and C.W. Hogue. BIND – a data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics*, 16:465–477, 2000.

[3] K. Ball, T.G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann. of Appl. Probab.*, 16(4):1925–1961, 2006.

[4] G. Batt, M. Page, I. Cantone, G. Goessler, P. Monteiro, and H. de Jong. Efficient parameter search for qualitative models of regulatory networks using symbolic model checking. *Bioinformatics*, 26(18):603–610, 2010.

[5] S.A. Becker, A.M. Feist, M.L. Mo, G. Hannum, B.O. Palson, and M.J. Herrgard. Quantitative prediction of cellular metabolism with constraint based models: the COBRA toolbox. *Nat. Protoc.*, 2:727–738, 2007.

[6] S. Berthoumieux, M. Brilli, H. de Jong, D. Kahn, and E. Cinquemani. Identification of metabolic network models from incomplete high-throughput datasets. *Bioinformatics*, 17(13):186–195, 2011.

[7] J. Bruck and M. Gibson. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem*, 105:1876–1889, 2000.

[8] Y. Cao, L.R. Petzold, M. Rathinam, and D.T. Gillespie. The numerical stability of leaping methods for stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 121:12169–12178, 2004.

[9] M.P. Cary, G.D. Bader, and C. Sander. Pathway information for systems biology. *FEBS Lett*, 579:1815–1820, 2005.

[10] R. Caspi et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acid Res.*, 38:473–479, 2010.

[11] W.W. Chen, M. Niepel, and P.K. Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes and Development*, 24:1861–1875, 2010.

[12] L.H. Chu and B.S. Chen. Construction of a cancer-perturbed protein-protein interaction network for discovery of apoptosis drug targets. *BMC Systems Biology*, 2(56):1–17, 2008.

[13] H.Y. Chuang, E. Lee, Y.T. Liu, D. Lee, and T. Ideker. Network based classification of breast cancer metastasis. *Mol. Syst. Biology*, 3:140, 2007.

[14] E.M. Clarke, J.R. Faeder, C.J. Langmead, L.A. Harris, S.K. Jha, and A. Legay. Statistical model checking in biolab: Applications to the automated analysis of t-cell receptor signalling pathway. In *CMSB*, volume 5307 of *LNCS*, pages 231–250. Springer, 2008.

[15] V. Danos and C. Laneve. Formal molecular biology. *Theor. Comp. Science*, 325:69–110, 2004.

[16] H. de Jong. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.*, 9(1):67–103, 2002.

[17] H. de Jong. Vers la cellule virtuelle. *DocSciences*, 8:20–25, 2009.

[18] H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyser: qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–344, 1966.

[19] H. de Jong and M. Page. Search for steady states of piecewise-linear differential equation models of genetic regulatory networks. *IEEE/ACM Trans. on Comp. Biol. and Bioinformatics*, 5(2):508–522, 2008.

[20] H. de Jong and D. Ropers. Strategies for dealing with incomplete information in the modeling of molecular interaction networks. *Brief. in Bioinformatics*, 7(4):354–363, 2006.

[21] E. Demir et al. PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics*, 18:996–1003, 2002.

[22] E. Demir et al. The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28:935–942, 2010.

[23] F. Devun, L. Walter, J. Belliere, C. Cottet-Rousselle, X. Leverve, and E. Fontaine. Ubiquinone analogs: a mitochondrial permeability transition pore-dependent pathway to selective cell death. *PLoS One*, 5(7):e11792, 2010.

[24] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, R. Srivas, and B.O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA*, 104(6):1777–1782, 2007.

[25] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.*, 13:2475–2484, 2003.

[26] S. Elmore. Apoptosis: a review of programmed cell death. *Toxicol. Pathol.*, 35(4):495–516, 2007.

[27] S.N. Ethier and T.G. Kurtz. *Markov processes: characterization and convergence*. Wiley series in Probability and Statistics. Wiley, New-York, 2005.

[28] M. Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rat. Mech. Anal.*, 132:311–370, 1995.

[29] J. Fisher and T. Henzinger. Executable cell biology. *Nat. Biotechnol*, 25:1239–1249, 2007.

[30] O. Folger et al. Predicting selective drug targets in cancer through metabolic networks. *Molecular Systems Biol.*, 7(501):1–10, 2011.

[31] C. Frezza et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*, 477(7363):225–228, 2011.

[32] C.W. Gardiner. *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, Berlin, 2004.

[33] E. Gasteiger et al. ExPaSy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, 31:3784–3788, 2003.

[34] I.M. Ghobrial, T.E. Witzig, and A.A. Adjei. Targeting apoptosis pathways in cancer therapy. *CA Cancer J. Clin*, 55(3):178–194, 2005.

[35] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.

[36] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[37] D.T. Gillespie. A rigourous derivation of the chemical master equation. *Physica A*, 188(1-3):405–425, 1992.

[38] D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716–1731, 2001.

[39] D.T. Gillespie, Y. Cao, and L. Petzold. Multicale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comp. Phys.*, 206(2):395–411, 2005.

[40] D.T. Gillespie, Y. Cao, and L. Petzold. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.*, 122:403–434, 2005.

[41] D.T. Gillespie, Y. Cao, K.R. Sanft, and L.R. Petzold. The subtle business of model reduction for stochastic chemical kinetics. *J. Chem. Phys.*, 130(6):064103, 2009.

[42] J. Goutsias. Quasi-equilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.*, 122(18):184102, 2005.

[43] J. Goutsias. Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophysical J.*, 92:2350–2365, 2007.

[44] J. Goutsias and N.H. Lee. Computational and experimental approaches for modeling gene regulatory networks. *Curr. Pharm. Design*, 13(14):1415–1436, 2007.

[45] F. Grognard, H. de Jong, and J.L. Gouzé. *Piecewise-linear models of genetic regulatory networks: theory and example*, pages 137–159. Number 357 in LN in Control and Inf. Sci. Springer, 2007.

[46] J. Gunawardena. A linear elimination framework for nonlinear biochemical systems. *Submitted*, 2011.

[47] R.N. Gutenkunst, J.J. Waterfall, F.P. Casey, K.S. Brown, C.R. Myers, and J.P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, 3:1871–1878, 2007.

[48] E.L. Haseltine and J.B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117:6959–6969, 2002.

[49] D.J. Higham. Modeling and simulating chemical reactions. *SIAM Rev.*, 50(2):347–368, 2008.

[50] W. Huang, B.T. Sherman, and R.A. Lampicki. Bioinformatics enrichment tolls: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, 37:1–17, 2009.

[51] H.J. Huber, H. Duessmann, J. Wenus, S.M. Kilbride, and J.H. Prehn. Mathematical modelling of the mitochondrial apoptosis pathway. *Biochim. Biophys. Acata*, 1813(4):608–615, 2011.

[52] M. Hucka et al. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *IEEE Systems Biology*, 1(1):41–53, 2004.

[53] K. Jaqaman and G. Danuser. Linking data to models: data regression. *Nat. Rev. Mol. Cell Biol.*, 7:1–10, Nov 2006.

[54] K.A. Johnson and R.S. Goody. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. *Biochemistry*, 50(4):8264–8269, 2011.

[55] M. Kanehisha, S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. The KEGG resource for deciphering the genome. *Nucleic Acid Res.*, 32:270–280, 2004.

[56] G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Rev. Mol. Cell Biol.*, 9(4):770–780, 2008.

[57] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.

[58] H. Kitano. A graphical notation for biochemical networks. *BioSilico*, 1:169–176, 2003.

[59] P.E. Kloeden, E. Platen, and H. Schurz. *Numerical solution of SDE through computer experiment*. Springer-Verlag, New York, 1997.

[60] V.N. Kolokoltsov. *Nonlinear Markov processes and kinetic Equations*, volume 182 of *Cambridge tracts in Mathematics*. Cambridge University Press, 2010.

[61] M. Krull et al. TRANSPATH: an information resource for storing and vizualing pathways and their pathological aberrations. *Nucleic Acid Res.*, 34:546–551, 2006.

[62] M. Krummenacker, S. Paley, L. Mueller, T. Yan, and P.D. Karp. Querying and computing with BioCyc databases. *Bioinformatics*, 21:3454–3455, 2005.

[63] S. Legewie, N. Blüthgen, and H. Herzel. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comp. Biol.*, 2(9):e120, 2006.

[64] C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M.I. Stefan, J.L. Snoep, M. Hucka, N. Le Novère, and C. Laibe. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4:92, Jun 2010.

[65] H. Li, Y. Cao, L. Petzold, and D.T Gillespie. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotechnology Progress*, 24:56–61, 2008.

[66] J. Liang and H. Qian. Computational cellular dynamics based on the cheical master equation: a challenge for understanding complexity. *J. Comp. Sci. Technol.*, 25(1):154–168, 2010.

[67] W.E.D. Liu and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetics systems with disparate rates. *J. Chem. Phys.*, 123(19):194107, 2002.

[68] L. Lok and R. Brent. Automatic generation of cellular reaction networks with MOLECULIZER 1.0. *Nat. Biotechnol.*, 23:131–136, 2011.

[69] A. Mallavarapu, M. Thomson, B. Ullian, and J. Gunawardena. Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *J. R. Soc. Interface*, 6:257–270, 2009.

[70] J.P. Massar, M. Travers, M. Elhai, and J. Shrager. BIOLINGUA: a programmable knowledge environment for biologists. *Bioinformatics*, 21:199–207, 2005.

[71] L. Matthews et al. Reactome knowledgebase of human biological pathways and processes. *Nucleic Acid Res.*, 37:619–622, 2009.

[72] J.M. McColluma, G.D. Peterson, C.D. Cox, M.L. Simpson, and N.F. Smatova. The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior. *J. Comp. Biol. Chem.*, 30:39–49, 2005.

[73] S. McNamara, A.M. Bersani, K. Burrage, and R.B. Sidje. Stochastic chemical kinetics and the total quasi-steady-state assumption: Application to the stochastic simulation algorithm and chemical master equation. *J. Chem. Phys.*, 129(9):095105, 2008.

[74] D.A. McQuarrie. Stochastic approach to chemical kinetics. *J. Appl. Probab.*, 4(3):413–478, 1967.

[75] L. Menten and M.I. Michaelis. Die Kinetic der Invertinwirkung. *Biochem. Z.*, 49:333–369, 1913.

[76] H. Mi et al. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.*, 33:284–288, 2005.

[77] H. Miao, X. Xia, A.S. Perelson, and H. Wu. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.*, 53(1):3–39, 2011.

[78] A.K. Miller et al. An overview of the CellML API and its implementations. *BMC Bioinformatics*, 11:178, 2010.

[79] P.T. Monteiro, E. Dumas, B. Besson, R. Mateescu, M. Page, A.T. Freitas, and H. de Jong. A service-oriented architecture for integrating the modeling and formal verification of genetic regulatory networks. *BMC Bioinformatics*, 10:450, 2009.

[80] E.C. O'Shaughnessy, S. Palani, J.J. Collins, and C.A. Sarkar. Tunable signal processing in synthetic MAP kinase cascades. *Cell*, 144:119–131, Jan 2011.

[81] H. Qian. The mathematical theory of molecular motor movement and chemomechanical energy transduction. *J. Math. Chem.*, 27(3):219–234, 2000.

[82] H. Qian and E.L. Elson. Single molecule enzymology: stochastic michelis-menten kinetics. *Biophysical Chem.*, 101–102:565–576, 2002.

[83] R. Ramaswamy, N. González-Segredo, and I.F. Sbalzarini. A new class of highly efficient exact simulation algorithms for chemical reaction networks. *J. Chem. Phys.*, 130(244104):1–13, 2009.

[84] R. Ramaswamy and I.F. Sbalzarini. A partial-propensity variant of the composition-rejection stochastic simulation algorithm for chemical reaction networks. *J. Chem. Phys.*, 132(044102):1–6, 2010.

[85] R. Ramaswamy and I.F. Sbalzarini. A partial-propensity formulation of the stochastic simulation algorithm for chemical rection networks with delays. *J. Chem. Phys.*, 134(014106):1–8, 2011.

[86] R. Ramaswamy, I.F. Sbalzarini, and N. González-Segredo. Noise-induced modulation of the relaxation kinetics around a non-equilibrium steady state of non-linear chemical reaction networks. *PLoS One*, 6(1):e16045, 2011.

[87] S. Ramsey, D. Orrel, and H. Bolouri. DIZZY: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.*, 3:415–436, 2005.

[88] C.V. Rao and A.P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J. Chem. Phys.*, 118(11):4999–5010, 2003.

[89] D. Ropers, V. Baldazzi, and H. de Jong. Model reduction using piecewise linear approximations preserves dynamic properties of the carbon starvation response in Escherichia Coli. *IEEE/ACM Trans. Comp. Biol. and Bioinformatics*, 8(1):166–181, 1966.

[90] S. Ryu, S.C. Lin, N. Ugel, M. Antoniotti, and B. Mishra. Mathematical modeling of the formation of apoptosome in intrinsic pathway of apoptosis. *Sys. Synth. Biol.*, 2(1-2):49–66, 2009.

[91] K.R. Sanft, D.T. Gillespie, and L.R. Petzold. Legitimacy of the stochastic michaelis-menten approximation. *IET Syst. Biol.*, 5(1):58–69, 2011.

[92] C.F. Schaefer et al. PID: the Pathway Interaction Database. *Nucleic Acid Res.*, 37:674–679, 2009.

[93] J. Schellenberger, J.O. Park, T.M. Conrad, and B.O. Palsson. BiGG: a biochemical genetic and genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11:213, 2010.

[94] M. Schulz, F. Krause, N. Le Novère, E. Klip, and W. Leibermeister. Retrieval, alignment and clustering of computational models based on semantic annotations. *Molecular Syst. Biology*, 5(512):1–10, 2011.

[95] V. Shahrezaei and P.S. Swain. The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19(4):369–374, 2008.

[96] P. Shannon et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 13:2498–2504, 2003.

[97] I. Shmulevitch and E.R. Dougherty. *Genomic signal processing.* Princeton series in Applied Mathematics. Princeton University Press, 2007.

[98] C.K. Speirs, M. Hwang, S. Kim, W. Li, S. Chang, V. Varki, L. Mitchell, S. Schleicher, and B. Lu. Harnessing the cell death pathway for targeted cancer treatment. *Am. J. Cancer Res.*, 1(1):43–61, 2011.

[99] M. Stiefenhofer. Quasi-steady-state approximation for chemical reaction networks. *J. Math. Biol.*, 36:593–609, 1998.

[100] I.W. Taylor et al. Dynamic modularity in interaction networks predicts breast cancer outcome. *Nature Biotechnology*, 27(2):199–204, 2009.

[101] M. Thomson and J. Gunawardena. The rational parametrization theorem for multisite post-translational modification systems. *J. Theor. Biol.*, 261:626–636, 2009.

[102] M. Thomson and J. Gunawardena. Unlimited multistability in multisite phophorylation systems. *Nature*, 460:274–277, 2009.

[103] N.N. van Kampen. *Stochastic processes in physics and chemistry.* North-Holland, Amsterdam, 1981.

[104] D.J. Wilkinson. *Stochastic modelling for systems biology.* Chapman & Hall/CRC, Boca Raton, FL, 2006.

[105] S. Wu, J. Fu, Y. Cao, and L.R. Petzold. Michelis-menten speeds up tau-leaping under a wide range of conditions. *J. Chem. Phys.*, 134:134112, 2011.

[106] Y. Yamamoto, K. Inoue, and A. Doncescu. Integrating abduction and induction in biological inference using CF-induction. In H. Lodhi and S. Muggleton, editors, *Elements of computational systems biology*, pages 213–234. Wiley, New York, 2010.

[107] N. Yosef et al. Toward accurate reconstruction of functional protein networks. *Molecular System Biology*, 5(248):1–19, 2009.

[108] G.U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. *Phil. Trans. Roy. Soc. London Ser. B*, 213:21–87, 1925.

[109] E.S. Zeron and M. Santillán. *Numerical solution of the chemical master equation: uniqueness and stability of the stationary distribution for chemical networks, and mRNA bursting in a gene network with negative feedback regulation*, volume 487 of *Methods in Enzymology*, chapter 6, pages 147–169. Elsevier, 2011.

[110] W. Zhou, X. Peng, Z. Yang, and Y. Wang. Accelerated stochastic simulation algorithm for coupled chemical reactions with delays. *Comput. Biol. Chem.*, 32:240–242, 2008.