An Alternative to Gillespie's Algorithm for Simulating Chemical Reactions

Roberto Barbuti, Andrea Maggiolo–Schettini,
Paolo Milazzo, and Angelo Troina

Dipartimento di Informatica, Università di Pisa
Largo B. Pontecorvo 3, 56127 - Pisa, Italy
{barbuti,maggiolo,milazzo,troina}@di.unipi.it

Abstract. We introduce a probabilistic algorithm for the simulation of chemical reactions, which can be used as an alternative to the well-established stochastic algorithm proposed by D.T. Gillespie in the '70s. We show that the probabilistic evolution of systems derived by means of our algorithm can be compared to the stochastic time evolution of chemical reactive systems described by Gillespie. Moreover, we use our algorithm in the definition of a formal model based on multiset rewriting, and we show some simulation results of enzymatic activity, which we compare with results of real experiments.

1 Introduction

The modelling of chemical reactions by using deterministic rate laws has been successful in both chemistry and biochemistry. The law of mass action is an empirical law which gives a relation between reaction rates and molecular concentrations. Given an initial concentration, we can use that law to derive all the concentrations at future time points.

The law is well tailored for describing deterministic and continuous processes, but it is less applicable when one considers smaller and smaller systems in which the discrete character of components supports the application of discrete models. Many of these models aim at describing the single steps of the evolution of a system [1–3]. Moreover, the inherent random character of chemical reactions justifies the use of discrete stochastic models [4–9]. Most of these models construct on the paper by Gillespie [10], which gives a stochastic simulation algorithm physically and mathematically well grounded from a kinetic theory point of view and which has had substantial improvements from the computational point of view [11–13].

Gillespie's approach simulates the time evolution of a chemically reacting system by determining when the next reaction will occur and what kind of reaction it will be. The time of the next reaction is individuated on the continuous time axis, thus the algorithm never approximates infinitesimal time movements by finite time steps. Kind and time of the next reaction are computed on the basis of a stochastic reaction constant. Such a constant is in general unknown

and it is guessed starting from the deterministic kinetic rate constant in an approximated way [7].

The main features of the method we propose in this paper are that time advances by discrete steps and that the reaction which will occur is computed directly from the deterministic kinetic rate. The use of the kinetic rate has the advantage of immediately relating real experiments and simulations. Actually, the reaction rates given by a real experiment can be used for a simulation, and vice–versa one may perform simulations with different reaction rates until a value is found that gives the observed behaviors. As a consequence of our choice of the reaction rate as the stochastic constant, our method does not depend on the measure unit chosen for reactants (number of molecules, moles, micromoles, picomoles, etc ...), and therefore it is better scalable than Gillespie's method.

We show that Gillespie's approach and ours are comparable under suitable assumptions. Moreover, we have implemented our method and applied it to samples proposed by Gillespie, obtaining the same results. We have also studied a real case of enzymatic activity, namely the reactions in the calf eye due to enzyme Sorbitol Dehydrogenase. Simulation results agree with the experimental ones both from the qualitative and the quantitative points of view. In [14], where we used a different choice criterion of the next reaction, the agreement with experiments was not as good from the quantitative point of view.

The algorithm for computing the possible evolutions allows the construction of a probabilistic transition system on which properties of the described system can be checked by a probabilistic model checker, such as, for example, PRISM [15]. To this purpose, we formalise the steps of the algorithm as derivation steps of a probabilistic rewriting system.

In Section 2, after giving some background, we recall Gillespie's algorithm, we give our algorithm and compare them. In Section 3 we define the Probabilistic MultiSet Rewriting as an example of formal model in which our algorithm is used. In Section 4 we show some results of simulation of the Lotka and of the Brussellator reactions, which can be compared with the results obtained by Gillespie. Moreover we show some results of simulation of the Sorbytol Dehydrogenase, which we compare with experimental results supplied by the authors of [16]. Finally, in Section 5 we give some conclusions.

2 Simulation of Chemical Reactions

2.1 Background

The fundamental empirical law governing reaction rates in biochemistry is the *law of mass action*. This states that for a reaction in a homogeneous medium, the reaction rate will be proportional to the concentrations of the individual reactants involved. A chemical reaction is usually represented by the following notation:

$$\ell_1 S_1 + \ell_2 S_2 \underset{k_{-1}}{\overset{k}{\rightleftharpoons}} \ell_3 S_3 + \ell_4 S_4$$

where $S_1, \ldots, S_4$ are molecules, $\ell_1, \ldots, \ell_4$ are their stoichiometric coefficients, and k, k_{-1} are the kinetic constants. We denote with L the sum of the stoichiometric coefficients, that is the total number of reactant molecules. The use of the symbol $\rightleftharpoons$ denotes that the reaction is *reversible* (i.e. it can occur in both directions). Irreversible reactions are denoted by the single arrow $\rightarrow$.

For example, given the simple reaction

$$2A \underset{k_{-1}}{\overset{k}{\rightleftharpoons}} B$$

the rate of the production of molecule B for the law of mass action is:

$$\frac{dB_+}{dt} = k[A]^2$$

and the rate of destruction of B is:

$$\frac{dB_-}{dt} = k_{-1}[B]$$

where $[A], [B]$ are the *concentrations* (i.e. moles over volume unit) of the respective molecules. In general, the rate of a reaction is:

$$k[S_1]^{\ell_1} \cdots [S_\rho]^{\ell_\rho}$$

where $S_1, \ldots, S_\rho$ are all the distinct molecular reactants of the reaction.

The rate of a reaction is usually expressed in $moles \cdot s^{-1}$ (it is a speed), therefore the measure unit of the kinetic constant is $moles^{-(L-1)} \cdot s^{-1}$.

2.2 Gillespie's Stochastic Algorithm

In [10] Gillespie gives a stochastic formulation of chemical kinetics that is based on the theory of collisions and that assumes a stochastic reaction constant c_μ for each considered chemical reaction R_μ. The reaction constant c_μ is such that $c_\mu dt$ is the probability that a particular combination of reactant molecules of R_μ will react in an infinitesimal time interval dt.

The probability that a reaction R_μ will occur in the whole solution in the time interval dt is given by $c_\mu dt$ multiplied by the number of distinct R_μ molecular reactant combinations. For instance, the reaction

$$R_1 : S_1 + S_2 \rightarrow 2S_1 \tag{1}$$

will occur in a solution with X_1 molecules S_1 and X_2 molecules S_2 with probability $X_1 X_2 c_1 dt$. Instead, the inverse reaction

$$R_2 : 2S_1 \rightarrow S_1 + S_2 \tag{2}$$

will occur with probability $\frac{X_1(X_1-1)}{2!}c_2 dt$. The number of distinct R_μ molecular reactant combinations is denoted by Gillespie with h_μ, hence, the probability of R_μ to occur in dt (denoted with $a_\mu dt$) is

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

Now, assuming that $S_1, \ldots, S_n$ are the only molecules that may appear in a chemical solution, a *state* of the simulation is a tuple $(X_1, \ldots, X_n)$ representing a solution containing X_i molecules S_i for each i in $1, \ldots, n$. Given a state $(X_1, \ldots, X_n)$, a set of reactions $R_1, \ldots, R_M$, and a value t representing the current time, the algorithm of Gillespie performs two steps:

1. The time $t+\tau$ at which the next reaction will occur is randomly chosen with τ exponentially distributed with parameter $\sum_{\nu=1}^{M} a_\nu$;
2. The reaction R_μ that has to occur at time $t+\tau$ is randomly chosen with probability $a_\mu dt$.

The function $P_g(\tau, \mu)dt$ represents the probability that the next reaction will occur in the solution in the infinitesimal time interval $(t+\tau, t+\tau+dt)$ and will be R_μ. The two steps of the algorithm imply

$$P_g(\tau, \mu)dt = P_g^0(\tau) \cdot a_\mu dt$$

where $P_g^0(\tau)$ corresponds to the probability that no reaction occurs in the time interval $(t, t+\tau)$. Since $P_g^0(\tau)$ is defined as

$$P_g^0(\tau) = \exp\left(-\sum_{\nu=1}^{M} a_\nu \tau\right)$$

we have, for $0 \leq \tau < \infty$,

$$P_g(\tau, \mu)dt = \exp\left(-\sum_{\nu=1}^{M} a_\nu \tau\right) \cdot a_\mu dt$$

Finally, the two steps of the algorithm can be implemented in accordance with $P_g(\tau, \mu)$ by choosing τ and μ as follows:

$$\tau = \left(\frac{1}{\sum_{\nu=1}^{M} a_\nu}\right) \ln\left(\frac{1}{r_1}\right) \qquad \mu = \text{the integer for which } \sum_{\nu=1}^{\mu-1} a_\nu < r_2 \sum_{\nu=1}^{M} a_\nu \leq \sum_{\nu=1}^{\mu} a_\nu$$

where $r_1, r_2 \in [0, 1]$ are two real values generated by a random number generator. After the execution of the two steps, the clock has to be updated to $t+\tau$ and the state has to be modified by subtracting the molecular reactants and adding the molecular products of R_μ.

2.3 The Probabilistic Simulation Algorithm

The probabilistic algorithm we propose in this paper as an alternative to Gillespie's method assumes that in a very small fixed time interval Δt at most one reaction may occur in a chemical solution. The length of Δt depends on the number and on the speeds of the chemical reactions.

As we reported in Section 2.1, the speed of a chemical reaction R_μ is given by $k_\mu [S_{\mu 1}]^{\ell_{\mu 1}} \cdots [S_{\mu \rho}]^{\ell_{\mu \rho}}$, where k_μ is the kinetic constant of R_μ, $[S_{\mu i}]$ is the concentration of the i–th molecular reactant and $\ell_{\mu i}$ is the stoichiometric constant of

$S_{\mu i}$ in R_μ. For the sake of simplicity, we assume concentrations to be expressed as the number of molecules per volume unit. Given a set of chemical reactions $R_1, \ldots, R_M$ and a volume V of the chemical solution, Δt has to be fixed to $\frac{1}{MN}$ seconds, where N is an arbitrarily great integer value such that

$$0 < \frac{V k_\mu [S_{\mu 1}]^{\ell_{\mu 1}} \cdots [S_{\mu \rho}]^{\ell_{\mu \rho}}}{N} \leq 1 \tag{3}$$

for $1 \leq \mu \leq M$ and for all the possible concentrations (assumed to be finite) of $S_{\mu 1}, \ldots, S_{\mu \rho}$. For instance, if we have the following two reactions:

$$\begin{aligned} &R_1 : S_1 + S_2 \rightarrow 2S_1 && \text{with } k_1 = 3 \text{ molecules}^{-1} s^{-1} \\ &R_2 : 2S_1 \rightarrow S_1 + S_2 && \text{with } k_2 = 2 \text{ molecules}^{-1} s^{-1} \end{aligned}$$

and the initial solution is $10S_1$ in the volume of 1 litre, then the concentrations of S_1 and S_2 may vary between $([S_1] = 10, [S_2] = 0)$ and $([S_1] = 1, [S_2] = 9)$. The maximum of $k_1[S_1][S_2]$ is $3 \cdot 5 \cdot 5 = 75$ and the maximum of $k_2[S_1]^2$ is $2 \cdot 10^2 = 200$, hence N has to be greater or equal to 200, and consequently $\Delta t \leq \frac{1}{2 \cdot 200} = \frac{1}{400}$ seconds.

We remark that the choice of the value of N is critical. Actually, if we take for N the minimum value satisfying (3), we gain as regards efficiency but we may lose precision. This happens in particular when concentrations of reactants are low and the minimum value of N satisfying (3) gives a Δt too wide to assume that at most one reaction occurs in it. In such cases a greater value of N must be chosen.

Once chosen N, the *probabilistic reaction constant* p_μ of the chemical reaction R_μ is defined as follows:

$$p_\mu = \frac{V k_\mu}{N}$$

and the probability of R_μ to occur is:

$$P(R_\mu) = \begin{cases} p_\mu [S_{\mu 1}]^{\ell_{\mu 1}} \cdots [S_{\mu \rho}]^{\ell_{\mu \rho}} = \frac{p_\mu \prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu}} & \textit{if } R_\mu \textit{ can occur} \\ 0 & \textit{otherwise} \end{cases}$$

where $X_{\mu i}$ is the number of molecules $S_{\mu i}$ in the solution, for $1 \leq i \leq \rho$. We remark that reaction R_μ can occur if $\ell_{\mu i} \leq X_{\mu i}$ and that the requirement given in (3) ensures that $0 \leq P(R_\mu) \leq 1$.

As in Gillespie's algorithm, a state of the simulation is a tuple $(X_1, \ldots, X_n)$ representing a solution containing X_i molecules S_i, and $S_1, \ldots, S_n$ are all the possible molecules. Given a state $(X_1, \ldots, X_n)$ and a value t representing the current time, the probabilistic simulation algorithm consists in the iteration of the following two steps:

1. A reaction R_μ is randomly chosen (all the reactions are equiprobable);
2. The chosen R_μ is performed with probability $P(R_\mu)$.

Therefore, we have that the probability of choosing and performing reaction R_μ in the time interval $(t, t + \Delta t)$ is

$$P(\mu) = \frac{1}{M} P(R_\mu)$$

and the probability of performing no reactions in the same time interval is

$$P_0 = 1 - \sum_{\nu=1}^{M} \frac{1}{M} P(R_\nu)$$

The two steps of the algorithm can be implemented using a standard random number generator. At each iteration the current time t has to be incremented by Δt, and, if a reaction has occurred, the state $(X_1, \ldots, X_n)$ has to be updated by subtracting the molecular reactants and by adding the molecular products of the reaction performed.

2.4 Comparing the Two Algorithms

We begin the comparison of the two algorithms by discussing some problems related to the existence of the reaction constant c_μ and the computation of the number of distinct molecular reactants combinations h_μ.

The reaction constant c_μ is in general unknown and it is usually estimated from the more familiar kinetic constant k_μ. As in the previous section, without loss of precision, we assume the concentration of a substance in a solution to be expressed as the number of molecules per volume unit, better than as the number of moles per volume unit. Hence, k_μ is expressed as $molecules^{-(L-1)}s^{-1}$ instead of $moles^{-(L-1)}s^{-1}$, where L is the total number of molecular reactants of R_μ. This assumption allows avoiding the Avogadro number in the formula:

$$c_\mu = \frac{k_\mu \prod_{i=1}^{\rho} \langle X_{\mu i} \rangle^{\ell_{\mu i}}}{\langle h_\mu \rangle V^{L_\mu - 1}} \tag{4}$$

where V is the volume of the chemical solution, ρ is the number of distinct molecular reactants of R_μ, $\langle X_{\mu i} \rangle$ is the average number of molecules $S_{\mu i}$ that may appear in a solution with $1 \leq i \leq \rho$, $\langle h_\mu \rangle$ is the average of h_μ, $\ell_{\mu i}$ is the stoichiometric constant of $S_{\mu i}$ and L_μ is $\ell_{\mu 1} + \cdots + \ell_{\mu \rho}$.

The number of distinct reactant combinations h_μ corresponds to a product of binomial coefficients. As shown in [7], assuming that the numbers of the molecules in the solution are large, we have that h_μ can be approximated as follows:

$$h_\mu \approx \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{\prod_{i=1}^{\rho} \ell_{\mu i}!} \tag{5}$$

Hence, Eq. (4) can be simplified as follows:

$$c_\mu \approx \frac{k_\mu \prod_{i=1}^{\rho} \ell_{\mu i}!}{V^{L_\mu - 1}} \tag{6}$$

For instance, for the reaction R_1 of (1) we have $h_1 \approx X_1 X_2$, therefore $c_1 \approx \frac{k_1}{V}$, and for the reaction R_2 of (2) we have $h_2 \approx \frac{(X_1)^2}{2!}$, and therefore, $c_2 \approx \frac{k_2 2!}{V}$.

Now, we show that if we consider an infinitesimal time interval dt instead of Δt, we obtain that the probabilities derived by our algorithm correspond to the probabilities of Gillespie's one, modulo the approximations described above. In particular, in the following lemma we show that the probability of a reaction R_μ to occur in the infinitesimal time interval $(t, t+dt)$ in Gillespie's stochastic approach is approximated by the probability of the same reaction to be chosen and performed by the probabilistic simulation algorithm.

Lemma 1. *If $dt = \Delta t$, it holds $a_\mu dt \approx P(\mu)$.*

Proof. We prove only the non trivial case $P(\mu) \neq 0$. By definition of a_μ we have $a_\mu dt = c_\mu h_\mu dt$, and moreover:

$$\begin{aligned}
c_\mu h_\mu dt &\approx \frac{k_\mu \prod_{i=1}^{\rho} \ell_{\mu i}!}{V^{L_\mu - 1}} \cdot \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{\prod_{i=1}^{\rho} \ell_{\mu i}!} \cdot dt && \text{by equations (5) and (6)} \\
&= \frac{k_\mu V}{MN} \cdot \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu}} && \text{because } dt = \Delta t = \frac{1}{MN} \\
&= \frac{1}{M} p_\mu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu}} = P(\mu) && \text{by definitions of } p_\mu \text{ and } P(\mu) \quad \square
\end{aligned}$$

Since the probability of no reaction in a time interval of length at least τ is equivalent to the probability of no reaction in Δt multiplied by itself $\lceil \frac{\tau}{\Delta t} \rceil$ times, we define $P_0(\tau)$ as follows:

$$P_0(\tau) = P_0^{\lceil \frac{\tau}{\Delta t} \rceil}$$

Moreover, let $P(\tau, \mu)$ be the probability that at time t the next reaction will be R_μ and it will occur in the infinitesimal time interval $(t+\tau, t+\tau+dt)$. $P(\tau, \mu)$ is the product of the probability of the occurrence of no reactions in the time interval $(t, t+\tau)$ and the probability of choosing and performing R_μ, that is:

$$P(\tau, \mu) = P_0(\tau) \cdot P(\mu)$$

We prove that the reaction probability density function $P_g(\tau, \mu)dt$ of Gillespie is approximated by $P(\tau, \mu)$ if the interval Δt corresponds to the infinitesimal interval dt.

Theorem 2. *If $dt = \Delta t$ it holds $P_g(\tau, \mu)dt \approx P(\tau, \mu)$.*

Proof. By definition of $P_g(\tau, \mu)$ we have $P_g(\tau, \mu)dt = P_g^0(\tau) \cdot a_\mu dt$; moreover:

$$\begin{aligned}
P_g^0(\tau) = \exp\left(-\sum_{\nu=1}^{M} c_\nu h_\nu \tau\right) &\approx \exp\left(-\sum_{\nu=1}^{M} p_\nu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{M V^{L_\mu}} \frac{\tau}{dt}\right) \text{ by Lemma 1} \\
&= \exp\left(-\sum_{\nu=1}^{M} V k_\nu dt \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu}} \frac{\tau}{dt}\right) = \exp\left(-\sum_{\nu=1}^{M} V k_\nu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu - 1}} \tau\right)
\end{aligned}$$

By definition of $P(\tau,\mu)$ we have $P(\tau,\mu) = P^0(\tau) \cdot P(\mu)$; moreover:

$$P_0(\tau) = \left(1 - \sum_{\nu=1}^{M} p_\nu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{MV^{L_\mu}}\right)^{\lceil \frac{\tau}{\Delta t} \rceil} = \left(1 - \sum_{\nu=1}^{M} Vk_\nu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{MNV^{L_\mu}}\right)^{\lceil MN\tau \rceil}$$

By applying the substitution $\alpha = -\sum_{\nu=1}^{M} Vk_\nu \frac{\prod_{i=1}^{\rho} (X_{\mu i})^{\ell_{\mu i}}}{V^{L_\mu}}$, we obtain:

$$P_g(\tau,\mu)dt \approx \exp(\alpha\tau)\, a_\mu dt \qquad \text{and} \qquad P(\tau,\mu) = (1 + \frac{\alpha}{MN})^{\lceil MN\tau \rceil} P(\mu)$$

Since $dt = \Delta t = \frac{1}{MN}$ is infinitesimal (hence N is close to infinity), we can prove that $P_g(\tau,\mu)dt \approx P(\tau,\mu)$ by Lemma 1 and because $\lim_{x\to\infty}(1 + \frac{1}{x})^{\lceil x \rceil} = e$. □

3 Probabilistic MultiSet Rewriting

In [5] the application to molecular systems of the *Stochastic π-calculus* [4] is validated by the stochastic formulation of the kinetics of chemical reactions given by Gillespie. More precisely, in [5] a variant of the *Stochastic π-calculus* is presented in which Gillespie's algorithm is used in the semantics of the model.

Another example of use of Gillespie's algorithm is the definition of the *Stochastic MultiSet Rewriting (SMSR)* [17]: such a formalism is based on a set of rewriting rules (representing chemical reactions) that can be applied to multisets (representing solutions) with probabilities derived by that algorithm.

In this paper we use our algorithm to define the *Probabilistic MultiSet Rewriting (PMSR)*. As the name suggests, this model is similar to SMSR, and it uses our probabilistic algorithm instead of Gillespie's one. We choose to apply our method to MultiSet Rewriting because it seems to us a very natural representation of chemical systems, it has a simple semantics, and it permits us to use probabilities of the simulation algorithm as they are.

3.1 Probabilistic MultiSet Rewriting (PMSR)

As in SMSR, in PMSR multisets represent chemical solutions and rewriting rules represent reactions. We assume a possibly infinite set Σ of multiset elements (also called molecules) and we define probabilistic rewriting rules as follows.

Definition 3 (PMSR Rule). *A* Probabilistic MultiSet Rewriting rule *is a triple* (M_1, p, M_2) *where:*

- M_1 *and* M_2 *are two different multisets with elements in* Σ*;*
- $p \in]0,1]$ *is the* probabilistic constant *of the rule.*

A probabilistic rewriting rule (M_1, p, M_2) *can be denoted also with the more usual notation* $M_1 \rightarrow_p M_2$.

The application of a rule R to a multiset M produces a new multiset M' which is the result of replacing in M the left–hand side with the right–hand side of R. If the left–hand side of a rule is not a sub-multiset of M, then such a rule cannot be applied.

Since multisets are transformed by rule applications as solutions are transformed by reactions, we can give as probabilities to rules the probabilities of the reactions, which are derived by our algorithm. Therefore, we have that the probabilistic constant p of a rule corresponds to the probabilistic reaction constant p_μ of the reaction R_μ associated to that rule. Now, a possible evolution over time of a chemical solution can be represented by a sequence of applications of rewriting rules starting from an initial multiset. We define PMSR systems as states of the evolution of a solution.

Definition 4 (PMSR System). *A* Probabilistic MultiSet Rewriting system *is a pair $(M, \mathcal{R})$, where M is a multiset of elements in Σ and $\mathcal{R}$ is a finite set of rewriting rules.*

We give the semantics of PMSR systems as a probabilistic transition system. Probabilities of the transitions are the same as in Sect. 2.3.

Definition 5 (Semantics). *The semantics of PMSR is the probabilistic transition system in which states are PMSR systems and transitions are described by the following inference rules:*

$$\frac{R_\mu \in \mathcal{R} \quad R_\mu = M_\mu \to_{p_\mu} M'_\mu \quad M_\mu \subseteq M}{(M, \mathcal{R}) \xrightarrow{P(\mu)} ((M \setminus M_\mu) \cup M'_\mu, \mathcal{R})} \qquad \frac{}{(M, \mathcal{R}) \xrightarrow{P_0} (M, \mathcal{R})}$$

It is intended that all the transitions consume a fixed amount of time Δt, and therefore one might count the time taken for performing a sequence of transitions. Since the semantics of the model is a probabilistic transition system, properties of the described system can be verified by means of a probabilistic model checker.

4 Applications

We consider Lotka and Brussellator reactions simulated by Gillespie in [10]. Lotka and Brussellator reactions are the following:

$$\textbf{Lotka} \qquad Y_1 \xrightarrow{k_1} 2Y_1 \qquad Y_1 + Y_2 \xrightarrow{k_2} 2Y_2 \qquad Y_2 \xrightarrow{k_3} Z$$

where $k_1 = 10$, $k_2 = 0.01$ and $k_3 = 10$;

$$\textbf{Brussellator} \qquad X \xrightarrow{k_1} X + Y_1 \qquad Y_1 \xrightarrow{k_2} Y_2 \qquad 2Y_1 + Y_2 \xrightarrow{k_3} 3Y_1 \qquad Y_1 \xrightarrow{k_4} Z$$

where $k_1 = 5000$, $k_2 = 50$, $k_3 = 0.000025$ and $k_4 = 5$. The constants given by Gillespie have been transformed into kinetic constants by applying Eq. (6).

In Figure 1 we show some experimental results for Lotka and Brussellator reactions. As in [10], the initial solution for the Lotka simulation is given by

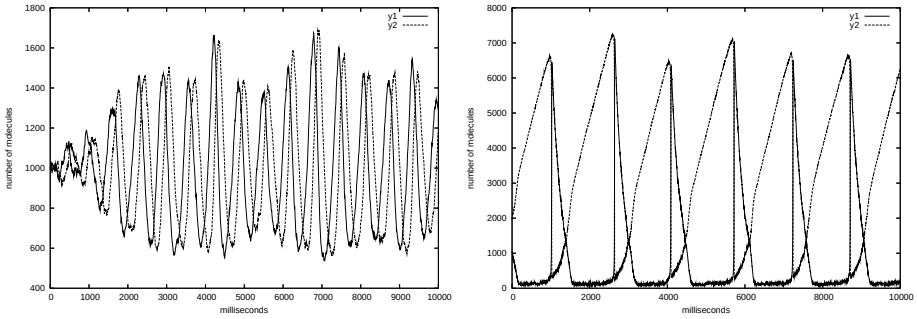

Fig. 1. Simulation results obtained with the Lotka reactions (on the left) and with the Brussellator reactions (on the right).

Setting	E	S	F	$NADH$	NAD^+	E–$NADH$	E–NAD^+
A	210	0	4×10^{11}	1.6×10^8	0	0	0
B	430	0	4×10^{11}	1.6×10^8	0	0	0

Table 1. Initial solutions. All values are in pM.

$Y_1 = Y_2 = 1000$, while for the Brussellator simulation we have an initial solution with $X = 1, Y_1 = 1000$ and $Y_2 = 2000$. Note that here the possible concentrations of reactants are infinite, and therefore N cannot be chosen such that (3) is satisfied for all the concentrations. Therefore we have chosen N such that (3) was satisfied for all the concentrations given by the steps of the simulations shown in Figure 1. For the Lotka simulation we used $N = 4 \cdot 10^4$, while for the Brussellator simulation we set $N = 2 \cdot 10^6$. Note that our results agree with those obtained by Gillespie with his stochastic algorithm.

As a further example, let us consider some reactions in the calf eye: here the enzyme Sorbitol Dehydrogenase (SDH) catalyses the reversible oxidation of Sorbitol and other polyalcohols to the corresponding keto–sugars (the accumulation of sorbitol in the calf eye has been proposed as the primary event in the development of sugar cataract in the calf [16]). The reactions are shown in the following scheme:

$$E + NADH \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E\text{–}NADH \qquad E \overset{k_7}{\rightarrow} E_i$$

$$E\text{–}NADH + F \underset{k_4}{\overset{k_3}{\rightleftharpoons}} E\text{–}NAD^+ + S \qquad E\text{–}NAD^+ \underset{k_6}{\overset{k_5}{\rightleftharpoons}} E + NAD^+$$

where E represents the enzyme Sorbitol dehydrogenase, S and F represent sorbitol and fructose, respectively, $NADH$ represents the nicotinamide adenine dinucleotide and NAD^+ is the oxidised form of $NADH$; $k_1, \ldots, k_7$ are the kinetic constants. Note that the enzyme degradation is modelled by the transformation of E into its inactive form E_i.

The kinetic constants are given in [16] (apart from k_7 that has been supplied by the authors of that paper) and are referred to concentrations measured in

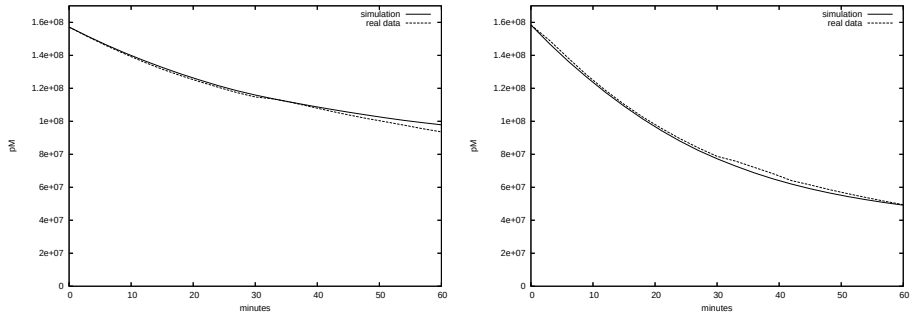

Fig. 2. Sorbitol dehydrogenase: concentrations of *NADH* with time varying. Simulations (solid lines) are compared with real experiments (dashed lines). The graph on the left corresponds to Setting A, while the graph on the right corresponds to Setting B.

$moles \cdot litres^{-1}$. The solutions used in experiments of biologists usually contain enzymes with a concentration in the order of $10^{-10} moles$ and other molecules with concentrations in the order of 10^{-1} up to $10^{-4} moles$. In order to be precise enough (but without considering interactions between single pairs of molecules) we adopt *picomoles* (equivalent to $10^{-12} moles$) as measure unit. We can use the measure unit we prefer because our algorithm does not depend on it. In general, by decreasing the measure unit we increase precision of the simulation but the values of the concentrations increase as well. Therefore we choose the scale such that the precision is sufficient still maintaining the values of the concentrations tractable. The kinetic constants for the considered reactions when this measure unit is *picomoles* are: $k_1 = 6.2 \times 10^{-6}$, $k_2 = 33$, $k_3 = 2.2 \times 10^{-9}$, $k_4 = 7.9 \times 10^{-9}$, $k_5 = 227$, $k_6 = 6.1 \times 10^{-7}$, and $k_7 = 1.9 \times 10^{-3}$. In Figure 2 we show the results given by the simulation with the initial solutions given by the settings A and B in Table 1, and we compare such results with the results of real chemical experiments. For setting A we set $N = 10^5$, while for setting B we used $N = 2 \cdot 10^5$.

The reactions due to Sorbitol Dehydrogenase have been studied in [14] with an algorithm that, as already mentioned, uses a different method for choosing rules. The results we obtained by the present algorithm are more satisfactory from the quantitative point of view. As one can see in the figures, the curves obtained by simulations practically coincide with the ones obtained by experiments.

5 Conclusions and Future Works

We have introduced a probabilistic algorithm for the simulation of chemical reactions that chooses the reaction occurring at every time on the basis of probabilities computed according to deterministic kinetic rates. We have compared our algorithm with the well–known Gillespie's algorithm and shown that results

of simulations agree both with experimental results and with Gillespie's simulations. Moreover, our method can be applied with different measure units, and, as a consequence, it may work when huge quantities of reactants are involved. In these cases the number of molecules is so big that Gillespie's algorithm, working at the level of molecules, is practically inapplicable.

References

1. Kam, N., Harel, D., Kugler, H., Marelly, R., Pnueli, A., Hubbard, E., Stern, M.: Formal modeling of c. elegans development: A scenario-based approach. In: Comp. Methods in Systems Biology 2003 (CMSB'03). Volume 2602 of LNCS. (2003) 4–20
2. Danos, V., Laneve, C.: Formal molecular biology. Theoretical Computer Science **325** (2004) 69–110
3. Curti, M., Degano, P., Baldari, C.: Causal π-calculus for biochemical modelling. In: Comp. Methods in Systems Biology 2003 (CMSB'03). Volume 2602 of LNCS. (2003) 21–33
4. Priami, C.: Stochastic π–calculus. The Computer Journal **38** (1995) 578–589
5. Priami, C., Regev, A., Shapiro, E., Silverman, W.: Application of a stochastic name-passing calculus to representation and simulation of molecular processes. Information Processing Letters **80** (2001) 25–31
6. Turner, T., Schnell, S., Burrage, K.: Stochastic approaches for modelling in vivo reactions. Computational Biology and Chemistry **28** (2004) 165–178
7. Wolkenhauer, O., Ullah, M., Kolch, W., Cho, K.H.: Modelling and simulation of intracellular dynamics: Choosing an appropriate framework. IEEE Transactions on NanoBioscience **3** (2004) 200–207
8. Nakanishi, T.: Stochastic analysis of an oscillating chemical reaction. Journal of the Physical Society of Japan **32** (1972) 1313–1322
9. Morton-Firth, C.J.: Stochastic simulation of cell signaling pathways. PhD thesis, University of Cambridge (1998)
10. Gillespie, D.: Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry **81** (1977) 2340–2361
11. Gibson, M., Bruck, J.: Efficient exact stochastic simulation of chemical systems with many species and many channels. J. of Phys. Chem. A **104** (2000) 1876–1889
12. Gillespie, D.: Approximate accelerated stochastic simulation of chemically reacting systems. Journal of Chemical Physics **115** (2001) 1716–1733
13. Bentele, M., Eils, R.: General stochastic hybrid method for the simulation of chemical reaction processes in cells. In: Comp. Methods in Systems Biology 2004 (CMSB'04). Volume 3082 of LNCS. (2005)
14. Barbuti, R., Cataudella, S., Maggiolo-Schettini, A., Milazzo, P., Troina, A.: A probabilistic model for molecular systems. To appear in Fundamenta Informaticae (2005). A preliminary version has appeared in the Proc. of the 13th Int. Workshop on Concurrency Specification and Programming (CS&P'04). Number 170 of Informatik–Berichte, Humboldt–Universitaet, Berlin, 202–216.
15. PRISM Model Checker: web site (2004) http://www.cs.bham.ac.uk/dxp/prism.
16. Marini, I., Bucchioni, L., Borella, P., Del Corso, A., Mura, U.: Sorbitol dehydrogenase from bovine lens: Purification and properties. Archives of Biochemistry and Biophysics **340** (1997) 383–391
17. Bistarelli, S., Cervesato, I., Lenzini, G., Marangoni, R., Martinelli, F.: On representing biological system through multiset rewriting. In: Proceedings of EUROCAST. (2003) 415–426