Journal of the Iranian Chemical Society, Vol. 3, No. 4, December 2006, pp. 327-333.

JOURNAL OF THE
Iranian
Chemical Society

Kinetic Investigation of Small Systems Using Different Algorithms

F. Taherkhani[a], G.A. Parsafar[a,*] and M.R. Rahimitabar[b]

[a]*Department of Chemistry and Nanotechnology Research Center, Sharif University of Technology, Tehran, Iran*

[b]*Department of Physics, Sharif University of Technology, Tehran, Iran*

(Received 10 June 2006, Accepted 8 August 2006)

We have investigated different algorithms for the simulation of kinetics of small systems. We have simulated the first order reversible reaction with the Gillespie, Gibson and Bruck time simulation as a function of Poisson distribution and compared the results of three algorithms. We have also simulated intracellular viral kinetics for a genome with Gillespie and Poisson distribution algorithms.

Keywords: Chemical kinetics, Small systems, Stochastic approach, Algorithms

INTRODUCTION

To investigate the kinetics of small systems, the use of a stochastic approach is necessary. The classical approach to chemical kinetics is called a deterministic approach, since once the state of the system is known at time t_1, its state at any other time will be known and no fluctuation about this value is observed [1,6].

The stochastic approach uses the inherent random nature of microscopic molecular collision to build a probabilistic model for a chemical reaction. This approach is useful in studying the kinetics of small systems. For small systems the validity of the deterministic approach becomes worse [1-7]. The calculated average concentration *vs*. time given by a stochastic approach and deterministic approach for a linear system are found to be equal. However, the results of these two approaches for nonlinear small systems are completely different [4].

The stochastic formulation proceeds by considering the grand probability function $P(X;t)$, the density probability of particles in volume V at time t, where X_i is the number of S_i species and $X \equiv (X_1, X_2 \ldots X_N)$ is a vector for the molecular species populations. Evidently, knowledge of this function provides a complete understanding of the probability distribution of all possible states at any time. By considering a discrete infinitesimal time interval $(t, t+dt)$ in which either 0 or 1 reaction occurs, we see that there exists only $M+1$ distinct configurations at time t that can lead to the state X at time $t+dt$. We can write our grand probability function at time $t+dt$ as a function of all possible states at time t as follows [2,9,12]:

$$P(X;t+dt) = P(X;t)*P \text{ (no state change over } dt) + \sum_{\mu=1}^{M} P(X - \upsilon_\mu;t)*P \text{ (state changes to } X \text{ over } dt)$$

where υ_μ is a stoichiometric vector defining the result of reaction μ on a state vector X, which means $X \longrightarrow X+\upsilon_\mu$ after an occurrence of reaction μ,P (no state changes over dt) = $1-\sum_{\mu=1}^{M} a_\mu(X)dt$, P (state changes to X over dt) = $\sum_{\mu=1}^{M} P(X-\upsilon_\mu;t)$.

By using this formula, we may derive the chemical master equation (CME) [2,9,12] that describes the stochastic dynamics of the system as

*Corresponding author. E-mail: parsafar@sharif.edu

$$\frac{\partial P(X;t)}{\partial t} = \sum_{\mu=1}^{M} a\mu(X - \upsilon_{\mu})*P(X - \upsilon_{\mu};t) - a_{\mu}(X)*P(X;t) \quad (1)$$

Due to the complexity of the CME, an analytical solution is rarely possible. Hence, for such a problem, simulation is an appropriate approach to solve the CME.

STOCHASTIC SIMULATION ALGORITHMS

Essentially, there are three modeling regimes, namely the discrete and stochastic, continuous and stochastic, and the continuous and deterministic, which depend on the nature of the reaction and the number of molecules in the system under study. A key simulation technique is the stochastic simulation approach to chemical reactions, which was developed by Gillespie *via* the stochastic simulation algorithm [8]. This is an exact procedure for numerically simulating the time evolution of a well stirred reacting system that takes proper account of the randomness inherent in such a system [2]. It is rigorously based on the same microphysical premise that underlies the CME described above and gives a more realistic representation of a system evolution than the deterministic reaction rate equation represented mathematically by ordinary differential equations that are to be solved simultaneously.

Recently, considerable attention has been paid to reduce the computational time of simulation algorithms for stochastic chemical kinetics. Gibson and Bruck [9] refined the first stochastic reaction simulation algorithm of Gillespie by reducing the number of random variables needed to be simulated. This algorithm can be effective for systems in which some reactions occur much more frequently than others. This algorithm is more efficient than Gillespie's direct method in the sense that only one new random number must be simulated for each reaction event that takes place, unlike Gillespie's method in which two random numbers are required. Note, however, that although selective recalculation of the hazards, $h_i(x,c_i)$ is also possible for the Gillespie algorithm [8,10,11], it could speed up the algorithm enormously for large systems, which will be introduced later.

Gillespie's Direct Method

For a system in a given state, Gillespie's direct algorithm [9] asks two equations: Which reaction occurs next and when does it occur? Clearly, both of these equations must be answered probabilistically by specifying the probability density $P(\mu,\tau)$ that the next reaction is μ and it occurs at time τ. It can be shown that

$$P(\mu,\tau)d\tau = a_{\mu}\exp(-\tau\sum_{j} a_j)d\tau \quad (2)$$

This equation leads directly to the answers of the two aforementioned questions. First, what is the probability distribution for reaction? Integrating $P(\mu,\tau)$ over all τ from 0 to ∞ results in

$$P(\mu) = \frac{a_{\mu}}{\sum_{j} a_j} \quad (3)$$

Second, what is the probability distribution as function of time? Summing $P(\mu,\tau)$ over all μ results in

$$P(\tau)d\tau = (\sum_{j} a_j)\exp(-\tau\sum_{j} a_j)d\tau \quad (4)$$

These two distributions lead to Gillespie's direct algorithm as:

1. Initialize (i.e., set initial numbers of molecules, set $0 \rightarrow t$) [9].
2. Calculate the propensity function, a_i, for all i.
3. Choose μ according to the distribution in equation 3.
4. Choose τ according to an exponential with parameter $\sum_{j} a_j$ (as in equation 4).
5. Change the number of molecules to reflect execution of reaction μ. Set $t + \tau \rightarrow t$
6. Go to step 2.

Gillespie's First Reaction Method

Algorithm I is direct in the sense that it generates μ and τ directly. Gillespie also developed the first reaction method, which generates a putative time τ_i for each reaction; the reaction with the smallest putative time is allowed to occur, which will be shown by τ_{μ}. Formally, the algorithm for the first reaction method is as follows [9]:

1. Initialize (i.e., set initial numbers of molecules, set $0 \longrightarrow t$).
2. Calculate the propensity function, a_i, for all i.
3. For each i, generate a putative time, τ_i, according to an exponential distribution with parameter a_i.
4. Let μ be the reaction whose putative time, τ_μ, is the smallest.
5. Let τ be τ_μ.
6. Change the number of molecules to reflect execution of reaction μ, set $t + \tau \longrightarrow t$.
7. Go to step 2 [9].

Gibson-Bruck Algorithm

The next reaction method (also known as the Gibson Bruck algorithm) is a modification of the first reaction method, which makes it much more efficient:

1. Initialize $0 \longrightarrow t$, in which c and x are the rate constant and number of particles in each time, and additionally calculate all of the initial reaction hazards $h_i(x,c_i)$, similar to the definition given in equation (9), $i = 1, 2, \ldots, n$. As shown in equations (5-8) use these hazards to simulate the first reaction time $t_i = [(1/h_i(x,c_i))\ln(1/r)]$ where r is a random number.

Suppose we have the following n reactions:

$$a_1 A_1 \xrightarrow{c_1} P_1 \tag{5}$$

$$a_2 A_2 \xrightarrow{c_2} P_2 \tag{6}$$

$$a_i A_i \xrightarrow{c_i} P_i \tag{7}$$

$$a_n A_n \xrightarrow{c_n} P_n \tag{8}$$

$h_i(A_i, c_i)$ at time t is defined as,

$$c_i \frac{k!}{a_i!(k-a_i)!}. \tag{9}$$

where k is number of A_i species at time t.
2. Let j be the index of the smallest t_i.
3. Set $t_j \longrightarrow t$.
4. Update x according to reaction with index j.
5. Update $h_j(x,c_j)$ according to the new state x and simulate a new putative time $t + [(1/h_j(x,c_j))\ln(1/r)] \rightarrow t_j$.
6. For each reaction $(i \neq j)$ whose hazard is changed by reaction j:
(a) Update $h_i' = h_i(x,c_i)$ (but temporarily keep the old h_i).
(b) Set $t + (h_i/h_i')(t_i - t) \longrightarrow t_i$
(c) Forget the old h_i.
7. If $t < T_{max}$, return to step 2 [9], where T_{max} is the input for the time length in which the reaction occurs.

We shall introduce a new algorithm based on the Poisson distribution $P(k) = e^{-\lambda}\lambda^k/k_1$ which was introduced in Ref. [2] but was not applied. This new algorithm is as follows:

1. Initialize the system at $t = 0$ with rate constants $c_1, c_2 \ldots c_n$ and initial numbers of molecules for each species, $A_1, A_2 \ldots, A_n$.
2. For each i, calculate $h_i(A_i,c_i)$ based on the current state.
3. Calculate $\lambda \equiv \sum_{i=1}^{n} h_i(A_i,c_i)$.
4. Simulate time to next event, t', as a *1/poisson(λ)* quantity.
5. Set $t = t + t'$.
6. Simulate reaction index, i, as a discrete random quantity with probabilities
$h_i(A_i,c_i)/\lambda$ and $i = 1, 2 \ldots, m$.
7. Update x according to reaction i'.
8. Output x and t.
9. If $t < t_{max}$, return to step 2.

This algorithm becomes more efficient by increasing the number of particles. Hence our new algorithm is better thanthe Gibson and Bruck algorithm. The time of simulation by the Poisson distribution for large systems is similar to the Gillespie algorithm.

ANALYSIS AND RESULTS

Exact Solution of Master Equation for First Order Reversible Reaction

If we let $X(t)$ be the concentration of A molecules at time t and let k_1 and k_2 be the forward and backward rate constants, respectively, then we obtain [7]

$$\frac{dP_x}{dt} = k_2(x_0 - x + 1)P_{x-1}(t) + k_1(x+1)P_{x+1}(t) - [k_1 x + k_2(x_0 - x)]P_x(t) \tag{10}$$

where is the total number of A and B molecules. By definition of generating function of $P_x(t)$ as

$$F(s,t)=\sum_{x=0}^{\infty}P_x(t)s^x \quad |s|<1 \tag{11}$$

Its partial differential equation becomes

$$\frac{\partial F}{\partial t}=[k_1+(k_2-k_1)s-k_1s^2]\frac{\partial F}{\partial s}+x_0k_2(s-1)F \tag{12}$$

If we assume that there are x_0 molecules of A at time zero, then the solution of equation (8) is

$$F(s,t)=[\frac{\lambda e^{-kt}(s-1)+\lambda-s}{\lambda}]^{x_0} \tag{13}$$

where $\lambda = k_1/k_2$ and $k = k_1 + k_2$. We can define the average and variance for any particle according to

$$< x >_t=\sum_{x=0}^{x_0}xP_x(t) \tag{14}$$

$$< x^2 >_t - < x >^2_t=\sum_{x=0}^{x_0}x^2P_x(t)-(\sum_{x=0}^{x_0}xP_x(t))^2 \tag{15}$$

According to equations (14) and (15) we can prove that,

$$E(X(t))=(\frac{\partial F}{\partial s})_{s=1} \tag{16}$$

$$D^2(X(t))=(\frac{\partial^2 F}{\partial s^2})_{s=1}+(\frac{\partial F}{\partial s})_{s=1}-(\frac{\partial F}{\partial s})^2_{s=1} \tag{17}$$

where $E(X(t))$ is the expected value or mean and $D^2(X(t))$ is the variance of $X(t)$.

We can use equations (16) and (17) to obtain the average number of particles and variance from equation (9) for a first order reversible reaction as

$$E(X(t))=[\frac{x_0}{(k_1+k_2)}](k_1e^{-kt}+k_2) \tag{18}$$

$$D^2(X(t))=[\frac{x_0\omega}{(1+\lambda)}](1-[\frac{\omega}{(1+\lambda)}]) \tag{19}$$

where $\omega = \lambda e^{-kt} + 1$.

Figures 1 and 2 show the simulation results of 500 runs of the stochastic algorithm simulating a system with the initial molecular populations A = 100, B = 2, for reaction $A \leftrightarrow B$ with k_f = 1 s^{-1}, k_r = 0.5 s^{-1} by the Gibson and Bruck and Gillespie algorithms, respectively. Figure 3 displays the simulation results obtained from the new algorithm, based on the Poisson distribution. In Figure 1, we observe that the average number of particles from the simulation is not the same as the average obtained from the master equation. Figure 2 is obtained from an exact procedure for numerically simulating the time evolution; however, the standard deviation about the mean value is high. The mean number of particles for simulation from the algorithm based on time simulation with the Poisson distribution is better than the number obtained from the prediction of the Gibson and Bruck algorithm, but the predicted mean for the number of particles from the Gillespie algorithm is better than that obtained from the Poisson distribution algorithm. In Figures 4, 5, and 6 the average number of particles obtained from the Gibson and Bruck and Gillespie time simulations are functions of the Poisson distribution algorithms, which are compared with the exact value. The variance about the mean for the Poisson distribution algorithm, as shown in Fig. 7, is less than those of the two other algorithms.

Intracellular Viral Model

We analyzed a simple network of a model virus, represented in equations (18-22). The components studied were the viral nucleic acids and a viral structural protein. The viral nucleic acids were classified as genomic (gen) or template (tem). The genome, whether it is DNA, positive-strand RNA, negative-strand RNA, or some other variant, is the vehicle by which viral genetic information is transported [3]. The genome can undergo one of the two fates. The first possibility is that it may be modified, whether through integration into the host genome or some other type of processing (reverse transcription) to form a template.

The template refers to the form of the nucleic acid that is transcribed and involved in catalytically synthesizing every viral component. The second possibility for the genome is that it may be packaged within structural proteins to form progeny virus structural proteins, such as capsid proteins, or envelope proteins [3].

The standard sequence of viral replication events involves

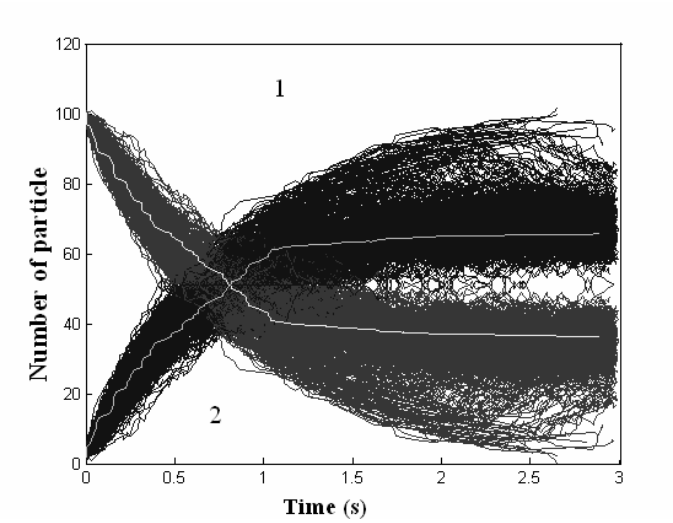

Fig. 1. Number and mean number of particles obtained from the Gibson and Bruck algorithm for first order reversible reactions A ↔ B when A = 100 and B = 2. Region 1 shows deviation from the mean number of particles, B, and region 2 shows deviation from the mean number of A.

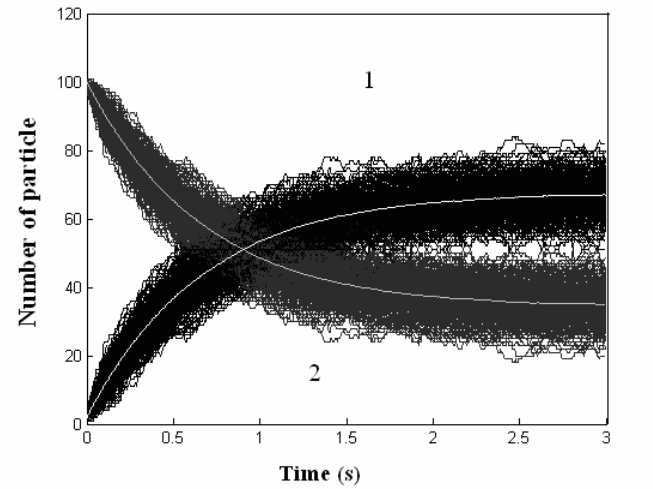

Fig. 2. Number and mean number of particles obtained from the Gillespie algorithm for first order reversible reactions A ↔ B when A = 100 and B = 2. Region 1 shows deviation from the mean Number of particles, B, and region 2 shows deviation from the mean number of A.

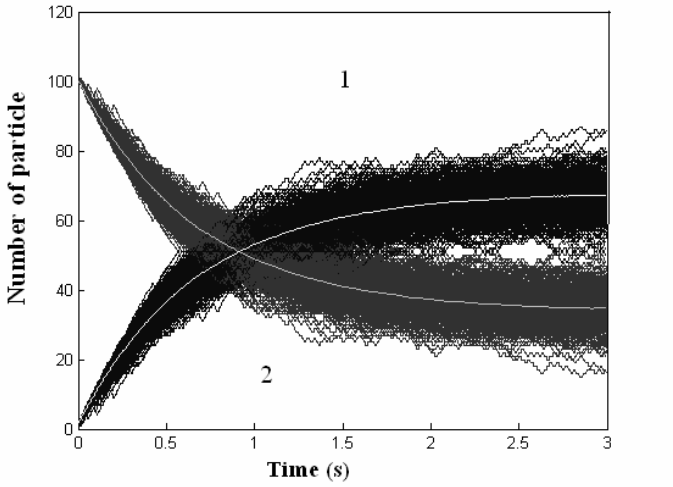

Fig. 3. Number and mean number of particles obtained for the time simulation as a function of Poisson distribution algorithm for first order reversible reactions A ↔ B when A = 100 and B = 2. Region 1 shows deviation from the mean number of particles, B, and region 2 shows deviation from the mean number of A.

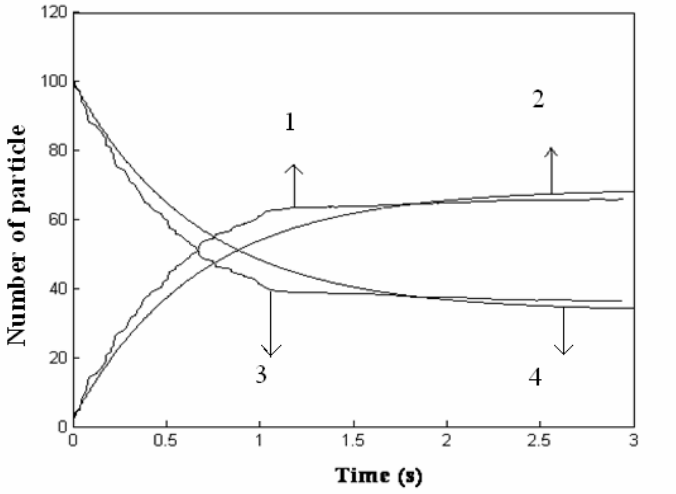

Fig. 4. Average number of particles obtained from Gibson and Bruck algorithm (1) compared to the exact value (2) for B. Similar result for A (3) is compared with the exact value (4).

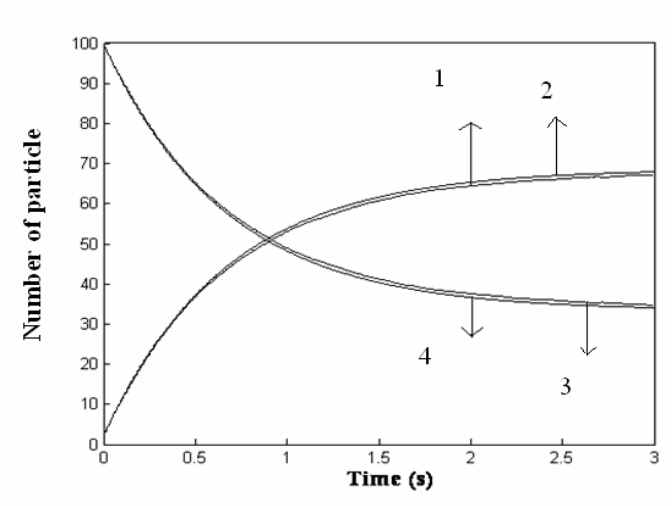

Fig. 5. Average number of particles obtained from the Gillespie algorithm (1) compared to the exact value (2) for B. The calculated result for A (3) is compared with the exact value (4).

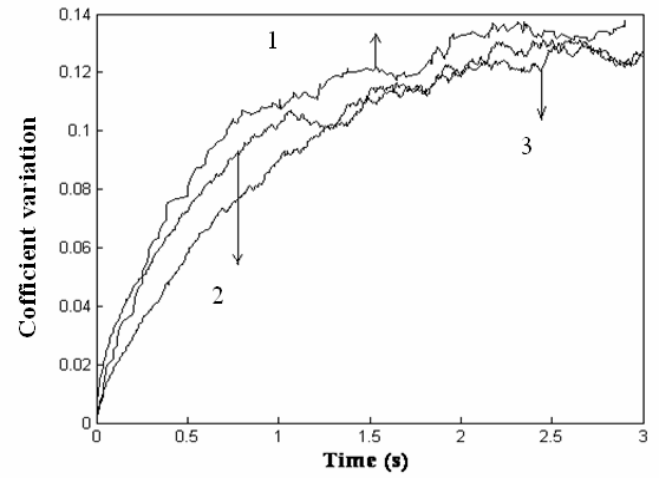

Fig. 7. Coefficient variation for the first order reversible reaction obtained from Bruck (1), Gillespie (2) and Poisson distribution (3) algorithms.

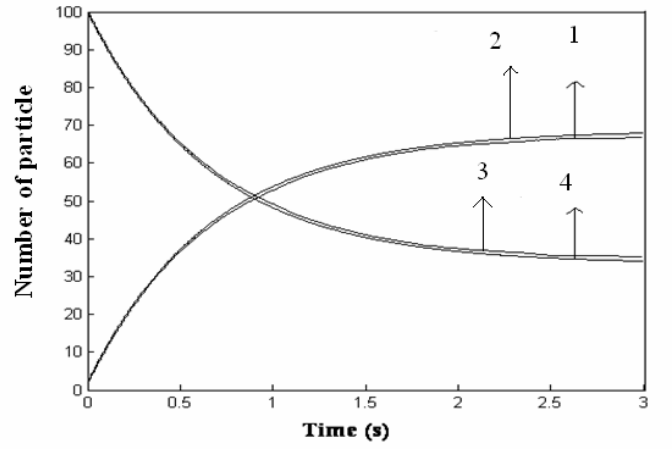

Fig. 6. Average number of particles obtained from the simulation as a function of Poisson distribution algorithm (1) compared to the exact value (2) for B. The calculated result for A (3) is compared with the exact value (4).

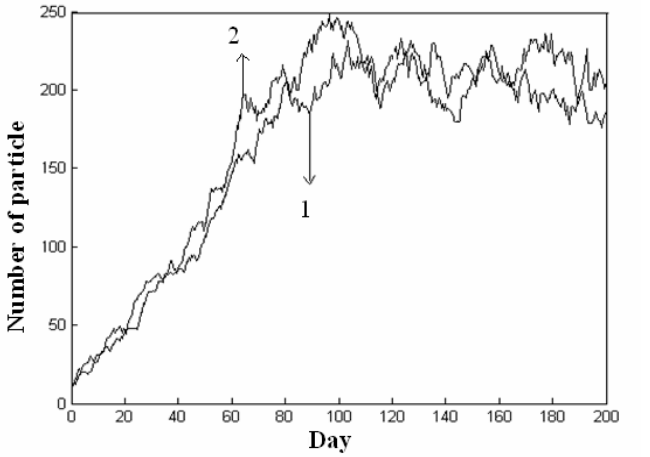

Fig. 8. Average number of particles for the genome specie in viral model obtained from Gillespie (1) and Poisson distribution (2) algorithms.

the amplification of the viral template after the infection, followed by production of progeny virus. DNA viruses, for example, initially make low levels of non-structural proteins, then catalytically amplify the number of template molecules to a level that is sufficiently high, so that structural proteins can be synthesized for incorporation into progeny particles. The

modeling network employed was "lumped", in that many individual reaction steps were combined into a single step. For example synthesis of structural protein requires that the viral DNA be transcribed into mRNA and that the mRNA be translated to generate the structural protein [3]. For the viral model, it was assumed that such reactions could be combined together and characterized using a single-rate parameter. The lumped reactions are represented by arrows with equations (18-19).

Figure 8 is one run from the genome simulation in an intracellular viral kinetics model with a time simulation using the Poisson distribution and Gillespie algorithms, which may represent the actual behavior of a reaction in a biological system [2]. The mechanism for the genome from the intracellular viral kinetics [3] is as follows:

$$A \underset{k_3}{\overset{k_1}{\rightleftharpoons}} X \tag{20}$$

$$X \xrightarrow{k_2} D \tag{21}$$

$$X \xrightarrow{k_5} B \tag{22}$$

$$B \xrightarrow{k_6} E \tag{23}$$

$$A+B \xrightarrow{k_4} C \tag{24}$$

In this mechanism A, B, C, D, E, and X are genome, structural protein, virus, degradation template, degradation structural protein, and template, respectively. The initial molecular populations are A = 10, B = 0, C = 0, D = 0, E = 0, X = 5 with $k_1 = 0.025$ day^{-1}, $k_2 = 0.25$ day^{-1}, $k_3 = 1.0$ day^{-1}, $k_4 = 7.5 \times 10^{-6}$ molecules^{-1} day^{-1}, $k_5 = 1000$ day^{-1}, $k_6 = 1.99$ day^{-1}.

DISCUSSION AND CONCLUSION

We simulated the concentration of a first order reversible reaction using three algorithms. We further simulated the intracellular viral kinetics of viral replication events by time simulation, as a function of the Poisson distribution and Gillespie algorithms. We have compared the mean number of particles given by the three algorithms with an exact solution of the master equation. Variance about the mean number of particles in the time algorithm based on the Poisson distribution $P(k) = e^{-\lambda}\lambda^k/k_1$, in which $t' = 1/Poisson(\lambda)$ quantity is less than those obtained from the two other algorithms. By increasing the number of particles, the mean number of particles approaches the exact average number of particles obtained from the master equation.

ACKNOWLEDGEMENTS

We appreciate Sharif University of Technology Research Council for its financial support and Mr. Mohammad Dehghan Niry for his useful comments.

REFERENCES

[1] D.A. McQuarrie, J. Chem. Phys. 38 (1963) 433.
[2] T.E. Turner, S. Schnell, K. Burrage, Com. Biol. & Chem. 28 (2004)165.
[3] R. Srivastava, L. You, J. Summers, J. Yin, J. Theor. Biol. 218 (2002) 309.
[4] J. Rose, Q. Zheng, J. Chem. Phys. 94 (1991) 3644.
[5] H. Qian, E.L.Elson, Biophys. Chem. 101 (2002) 565.
[6] D.A. McQuarrie, C.J. Jachimowski, M.E. Russell, J. Chem. Phys. 40 (1964) 2914.
[7] D.A. McQuarri, Stochastic Approach to Chemical Kinetics, Methuen, London, 1967.
[8] D.T. Gillespie, J. Phys. Chem.81 (1977) 2340.
[9] M.A. Gibson, J. Bruck, J. Phys. Chem. A 104 (2000) 1876.
[10] P. Erdi, J. Toth, Mathematical Models of Chemical Reaction. Theory and Applications of Deterministic and Stochastic Models, Princeton University Press, Princeton, NJ, 1988.
[11] D.T. Gillespie, J. Comput. Phys. 22 (1976) 403.
[12] D.T. Gillespie, J. Stat. Phys.16 (1977) 311.
[13] N.G. Van Kampen, Stochastic Processes in Physics and Chemistry, Elsevier, Amsterdam, 1992.