Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Discrete stochastic simulation of cell signaling: comparison of computational tools

Tiina Manninen*†*, Eeva Mäkiraatikka*, Antti Ylipää*, Antti Pettinen*, Kalle Leinonen* and Marja-Leena Linne**
*Institute of Signal Processing
†Institute of Mathematics
Tampere University of Technology
P.O. Box 553, FI-33101 Tampere, Finland
*Corresponding authors: tiina.manninen@tut.fi, marja-leena.linne@tut.fi

***Abstract*— Several stochastic simulation tools have been developed recently for cell signaling. A comparative evaluation of the stochastic simulation tools is needed to highlight the current state of the development. In our study, we have chosen to evaluate three stochastic simulation tools: Dizzy, Systems Biology Toolbox, and Copasi, using our own MATLAB® implementation as a benchmark. The Gillespie stochastic simulation algorithm is used in all tests. With all the tools, we are able to simulate stochastically the behavior of the selected test case and to produce similar results as our own MATLAB® implementation. However, it is not possible to use time-dependent inputs in stochastic simulations in Systems Biology Toolbox and Copasi. The present study is one of the first evaluations of stochastic simulation tools for realistic signal transduction pathways.**

***Index Terms*— Stochastic simulation tool, Gillespie stochastic simulation algorithm, signal transduction pathway**

I. INTRODUCTION

In recent years, the use of quantitative stochastic modeling and simulation has increased in biology. There is more and more evidence that deterministic approaches may not be adequate for modeling the complex reaction-diffusion systems often involving low numbers of signaling molecules [1], [2], [3]. Several stochastic approaches for simulating biochemical reactions have been developed over the last 30 years. The Gillespie stochastic simulation algorithm (SSA) [4], [5], the StochSim algorithm [6], among others, represent the discrete stochastic simulations. The chemical Langevin equation [7] and the chemical Fokker-Planck equation [8], for example, represent the continuous stochastic simulations. To facilitate the implementation of deterministic and stochastic approaches, many biochemical simulation approaches and tools have been presented, e.g. in [1], [2], [9], [10], [11].

Due to the fast development of the research area, there is a need for detailed comparative evaluation of the developed stochastic simulation tools for cell signaling. In our study, we have chosen to evaluate three stochastic simulation tools: Dizzy [12], Systems Biology Toolbox [13], and Copasi [14], and compare them to our own MATLAB® implementation, in which the SSA is used.

II. METHODS

A. Test case of cell signaling

We have recently studied a model of the protein kinase C (PKC) signal transduction pathway [10], [15], [16], [17] obtained from [18], [19]. The PKC pathway has been found to be important in several neuronal functions, including synaptic long-term potentiation (LTP) and depression (LTD). The signaling involved in LTP and LTD may form a biochemical basis for memory and learning [18], [20].

The PKC pathway model (Fig. 1) is obtained from the Database of Quantitative Cellular Signaling (DOQCS) [19]. The model describes the transformation of inactive PKC (PKCi) to active PKC (PKCa). The model has been originally presented for a hippocampal neuron [18] and it consists of 10 reversible reactions and 15 different interacting chemical species, of which 11 are model variables. Three of the model species correspond to second messengers and are used as model inputs: calcium ion (Ca^{2+}), arachidonic acid (AA), and diacylglycerol (DAG). The model has six computational intermediates which are summed as the output, PKCa.

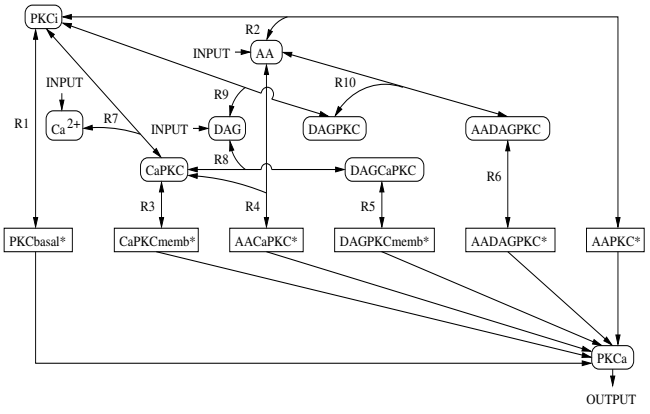

Fig. 1. Graphical design model of the protein kinase C pathway, originally published in Bioinformatics by Oxford University Press [10]. The original model is available in [19]. Reversible reactions R1 – R10 lead to formation of PKCa. Chemical species marked with asterisks (∗) are computational intermediates. Ca^{2+}, AA, and DAG are used as model inputs.

B. Gillespie stochastic simulation algorithm

The chemical master equation [4], [5] is often mathematically intractable. The well-established SSA is a systematic, computer-oriented procedure in which Monte Carlo techniques are employed to numerically simulate the discrete Markov process that the master equation describes analytically [4]. There are five ways to implement the SSA: using either the direct method [4], the first reaction method [4], the next reaction method [21], the optimized direct method [22], or the sorting direct method [23]. In the direct method [4], a random number r_2 is used to choose which reaction will occur, and another random number r_1 determines how long the step will last. The chemical species are altered according to the stoichiometry of the reaction ν_μ and the process is repeated. This detailed stochastic simulation algorithm is computationally expensive since only one reaction can take place in one time interval. With two random numbers r_1 and r_2, uniformly distributed on the interval $[0,1]$, the time interval $[t,t+\tau)$ is determined by the equation

$$\tau = \frac{1}{a_0} \ln \frac{1}{r_1}, \tag{1}$$

where a_0 is the sum of the propensity functions a_μ of each reaction occurring relative to another, $a_0 = \sum_{\mu=1}^{n} a_\mu$, and the reaction R$\mu$ is determined by the equation

$$\sum_{\mu'=1}^{\mu-1} a_{\mu'} < r_2 a_0 \leq \sum_{\mu'=1}^{\mu} a_{\mu'}. \tag{2}$$

The propensity functions a_μ are obtained by multiplying the rate constant of each reaction c_μ by the numbers of possible combinations of chemical species involved in each reaction h_μ, that is $a_\mu = c_\mu h_\mu$, where c_μ is in units of $1/\mathrm{s}$. For example, in the case of the reaction $\mathrm{X}_1 + \mathrm{X}_2 \rightarrow \mathrm{X}_3$, $a_1 = c_1 \mathrm{X}_1 \mathrm{X}_2$. The update of the system is then given by

$$\mathbf{X}(t+\tau) = \mathbf{X}(t) + \nu_\mu, \tag{3}$$

where $\mathbf{X}$ describes the numbers of chemical species.

III. STOCHASTIC SIMULATION TOOLS

Several stochastic simulation tools are available, e.g. Basis [24], BioNetGen [25], Cellware [26], Copasi [14], Dizzy [12], E-CELL [27], SimBiology [28], SmartCell [29], StochSim [6], STOCKS [30], and Systems Biology Toolbox [13]. In our study, we have chosen to evaluate three stochastic simulation tools, Dizzy, Systems Biology Toolbox, and Copasi, because both deterministic and stochastic simulations are available in these tools. Next, the three tools are evaluated according to their applicability, usability, and benefits and drawbacks. Test simulations are performed and compared to our own MATLAB® implementation. Our MATLAB® implementation is very flexible, e.g. time-dependent functions are possible as inputs and all data points can be saved as well as plotted.

IV. RESULTS

A. Applicability

Dizzy (Version 1.11.2) [12] is a simulation software package written in Java. There are four different solvers to simulate models deterministically, and four separate algorithms to simulate models stochastically: the two ways of the SSA (the next reaction method and the direct method), and the two types of Gillespie tau-leap algorithms, one for complex models and the other for simple ones. Dizzy also supports the Systems Biology Markup Language (SBML), and is capable of displaying models graphically. Dizzy runs on Microsoft Windows, UNIX, Linux, and Mac OS X platforms.

In Systems Biology Toolbox (Version 1.5) [13], the user can choose between seven deterministic solvers, and three stochastic algorithms: the SSA (the direct method), the binomial tau-leap algorithm, and the Poisson tau-leap algorithm. The toolbox supports the SBML and runs on Microsoft Windows, UNIX, and Linux platforms. However, stochastic simulation is only available on Microsoft Windows platform.

Copasi (Release Candidate 1, Build 17) has one solver to simulate models deterministically and one algorithm, the SSA (the direct method), to simulate models stochastically. Copasi supports both the SBML and loading of Gepasi [31] files. Copasi runs on Microsoft Windows, Linux, Mac OS X, and Solaris platforms.

B. Usability

The syntax used in Dizzy is simple. The manual includes step-by-step instructions how to get started with the software. For example, first the user can simulate ready-made models with very simple reactions and then extend the samples towards one's particular interest. In the implementation, the numbers of species defined in chemical reactions are always assumed to be dynamic. The inputs for chemical reactions are defined using a specific symbol in the syntax. Simulated responses can be plotted, listed as tables, or saved as csv-files.

Systems Biology Toolbox for MATLAB® provides a tutorial which is available in the Systems Biology Toolbox webpage [13]. The use of the editors (SBedit and SBeditBC) is uninformative, at least in the very beginning. The syntax used in Systems Biology Toolbox is more limited in the stochastic simulations than in the deterministic simulations. Simulated responses can be plotted or stored in MATLAB® structure arrays.

Copasi is used through a graphical user interface (GUI). Even though Copasi is still in early developmental stage, the GUI is very intuitive to use, especially if the user has experience with Gepasi, the predecessor of Copasi. The implementation of models is easy and straightforward, and the user can easily set up different tasks. For example, when the user sets up the reactions, Copasi automatically determines the reactants from the reactions. Copasi provides a simple tutorial and a manual, which cover most of the features. The simulated responses can be plotted and also saved as ASCII-files.

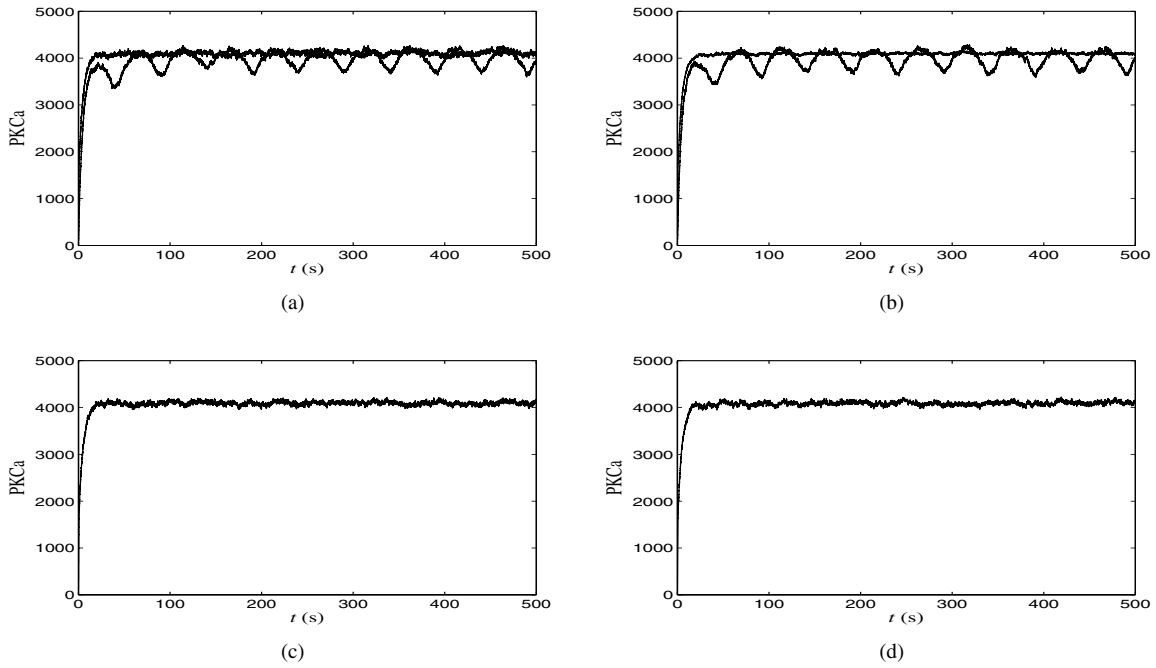

Fig. 2. Realizations of PKCa using stochastic simulation tools. V is 10^{-17} m^3 and initial number of PKCa is 120. (a) Own MATLAB® implementation. (b) Dizzy. (c) Copasi. (d) Systems Biology Toolbox. With Systems Biology Toolbox and Copasi, it is not possible to use a time-dependent sine wave input.

C. Benefits and drawbacks

The installation of Dizzy is very easy even for beginners. The error messages appearing during the implementation of models and the output plots are uninformative. The output plot contains no labels for the axes, no possibility to zoom plots, and there are symbols plotted around the result points. Furthermore, it is not possible to save all the data points into a file from simulations. There are no parameter estimation and sensitivity analysis available in Dizzy.

The installation of Systems Biology Toolbox is easy. The more the user is used to use MATLAB® the easier the use of the toolbox is. The error messages appearing during the use of the editors are not very informative. Furthermore, Model Parameters -object cannot include any calculation at all, e.g. the user needs to take care of the unit conversion. Furthermore, time-dependent functions cannot be used in stochastic simulations as inputs. Parameter estimation and parameter sensitivity analysis are available in Systems Biology Toolbox.

The installation of Copasi is very easy. Copasi offers several options in the model implementation, e.g. the user can use either concentrations or numbers of species. However, Copasi does not have a straightforward way of using inputs in models. With deterministic simulations, one can at least use a work-around solution similar to Gepasi [10], but with stochastic simulations the utilization of such a solution is not possible, at least to our knowledge. The error messages are not very informative for the user. For plotting, Copasi provides an useful output assistant. Furthermore, Copasi provides a good selection of additional tools, e.g. parameter estimation, sensitivity analysis, and stoichiometric analysis.

D. Simulation results

In the simulations, the Gillespie stochastic simulation algorithm (the direct method) is used. We have chosen to use a volume of 10^{-17} m^3. Two kinds of inputs for Ca^{2+}, AA, and DAG are used. First, all three are kept constant, i.e. the amounts of Ca^{2+}, AA, and DAG are 60221 (10 μM), 301107 (50 μM), and 903321 (150 μM), respectively. Second, the amounts of AA and DAG are kept as above, but the Ca^{2+} input follows a sine wave $(\sin(2\pi t/50)+1.1)\times 6022.1415$ $((\sin(2\pi t/50)+1.1)$ μM). It is not possible to use time-dependent functions in stochastic simulations in Systems Biology Toolbox and Copasi, and thus we show only simulations where constant inputs are used.

In Fig. 2, we show one realization for each input set for each simulation tool. The results show that realizations simulated using our own MATLAB® implementation, Dizzy, Systems Biology Toolbox, and Copasi quite accurately follow each other. With Systems Biology Toolbox and Copasi, it is only possible to use constant inputs in stochastic simulations.

V. CONCLUSIONS

Several stochastic simulation tools with similar capabilities have recently been developed. In order to guide the future development of simulation tools towards more user-friendly and versatile tools, a detailed comparative evaluation of existing tools is necessary. In the beginning of our study, we made a large survey of available stochastic simulation tools and chose to examine Dizzy, Systems Biology Toolbox, and Copasi in more detail.

In this study, we evaluate the tools according to their applicability, usability, documentation, analysis tools, and simulation results. The simulation results are benchmarked using our own MATLAB® implementation of the test case related to signal transduction. Based on our studies, the following issues can be concluded: 1) the usability of the simulation tool is crucial for the user, e.g. Dizzy is very easy to use even for beginners but Systems Biology Toolbox requires experience in MATLAB®; 2) proper manuals and analysis tools, as well as the possibility to easily exchange models between tools, e.g. using SBML, will be a necessity in the future; 3) the ability to utilize realistic external stimuli is very important especially in the case of realistic models (see also [10]), however, it is not possible in stochastic simulations in Copasi and Systems Biology Toolbox; and 4) Dizzy, Systems Biology Toolbox, and Copasi produce similar results as the benchmark. In summary, the present comparison is one of the first evaluations of stochastic simulation tools for realistic signal transduction pathways.

VI. ACKNOWLEDGMENTS

This work was financially supported by Tampere University of Technology Graduate School, and in part by the Academy of Finland, project nos. 213462 (Finnish Centre of Excellence program 2006 - 2011), 104508, 106030, and 107694, and by the Finnish Funding Agency for Technology and Innovation. The support of the Jenny and Antti Wihuri Foundation, the Ulla Tuominen Foundation, and the Foundation of Technology (for T.M.) is also acknowledged.

REFERENCES

[1] C. Lemerle, B. Di Ventura, and L. Serrano, "Space as the final frontier in stochastic simulations of biological systems," *FEBS Lett.*, vol. 579, pp. 1789–1794, 2005.

[2] C. M. Tan, S. Somani, and P. Dhar, "Modeling and simulation of biochemical systems with stochasticity," *In Silico Biol.*, vol. 4, no. 3, pp. 293–309, 2004.

[3] T. E. Turner, S. Schnell, and K. Burrage, "Stochastic approaches for modelling in vivo reactions," *Comput. Biol. Chem.*, vol. 28, pp. 165–178, 2004.

[4] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22, no. 4, pp. 403–434, 1976.

[5] ——, "Exact stochastic simulation of coupled chemical reactions," *J. Phys. Chem.*, vol. 81, no. 25, pp. 2340–2361, 1977.

[6] N. Le Novère and T. S. Shimizu, "StochSim: modelling of stochastic biomolecular processes," *Bioinformatics*, vol. 17, no. 6, pp. 575–576, 2001. [Online]. Available: http://www.pdn.cam.ac.uk/groups/comp-cell/StochSim.html

[7] D. T. Gillespie, "The chemical Langevin equation," *J. Chem. Phys.*, vol. 113, no. 1, pp. 297–306, 2000.

[8] ——, "Approximating the master equation by Fokker-Planck-type equations for single-variable chemical systems," *J. Phys. Chem.*, vol. 72, no. 10, pp. 5363–5370, 1980.

[9] A. P. Arkin, "Synthetic cell biology," *Curr. Opin. Biotechnol.*, vol. 12, pp. 638–644, 2001.

[10] A. Pettinen, T. Aho, O.-P. Smolander, T. Manninen, A. Saarinen, K.-L. Taattola, O. Yli-Harja, and M.-L. Linne, "Simulation tools for biochemical networks: evaluation of performance and usability," *Bioinformatics*, vol. 21, no. 3, pp. 357–363, 2005.

[11] B. M. Slepchenko, J. C. Schaff, J. H. Carson, and L. M. Loew, "Computational cell biology: spatiotemporal simulation of cellular events," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 31, pp. 423–441, 2002.

[12] S. Ramsey, D. Orrell, and H. Bolouri, "Dizzy: stochastic simulations of large-scale genetic regulatory networks," *J. Bioinform. and Comput. Biol.*, vol. 3, no. 2, pp. 415–436, 2005. [Online]. Available: http://magnet.systemsbiology.net/software/Dizzy/

[13] H. Schmidt and M. Jirstrand, "Systems Biology Toolbox for MATLAB: a computational platform for research in systems biology," *Bioinformatics*, vol. 22, no. 4, pp. 514–515, 2006. [Online]. Available: http://www.sbtoolbox.org/

[14] [Online]. Available: http://www.copasi.org/

[15] T. Manninen, M.-L. Linne, and K. Ruohonen, "Developing Itô stochastic differential equation models for neuronal signal transduction pathways," *Accepted in Computational Biology and Chemistry*, 2006.

[16] ——, "A novel approach to model neuronal signal transduction using stochastic differential equations," *Neurocomputing*, vol. 69, pp. 1066–1069, 2006.

[17] T. Manninen, A. Saarinen, K. Ruohonen, and M.-L. Linne, "Simulation study of deterministic differential equation model for protein kinase c signaling: sensitivity of stimuli, parameter values, and initial concentrations," in *Proceedings of the 4th TICSP Workshop on Computational Systems Biology, WCSB 2006*, P. Ruusuvuori, T. Manninen, H. Huttunen, M.-L. Linne, and O. Yli-Harja, Eds. Tampere University of Technology, Tampere, Finland, June 2006, pp. 61–64.

[18] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, pp. 381–387, 1999.

[19] S. Sivakumaran, S. Hariharaputran, J. Mishra, and U. S. Bhalla, "The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks," *Bioinformatics*, vol. 19, no. 3, pp. 408–415, 2003. [Online]. Available: http://doqcs.ncbs.res.in/

[20] T. V. P. Bliss and G. L. Collingridge, "A synaptic model of memory: long-term potentiation in the hippocampus," *Nature*, vol. 361, pp. 31–39, 1993.

[21] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *J. Phys. Chem. A*, vol. 104, pp. 1876–1889, 2000.

[22] Y. Cao, H. Li, and L. Petzold, "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems," *J. Chem. Phys.*, vol. 121, no. 9, pp. 4059–4067, 2004.

[23] J. M. McCollum, G. D. Peterson, C. D. Cox, M. L. Simpson, and N. F. Samatova, "The sorting direct method for stochastic simulation of biochemical systems with varying reaction execution behavior," *Comput. Biol. Chem.*, vol. 30, no. 1, pp. 39–49, 2006.

[24] T. B. L. Kirkwood, R. J. Boys, C. S. Gillespie, C. J. Proctor, D. P. Shanley, and D. J. Wilkinson, "Towards an e-biology of ageing: integrating theory and data," *Nat. Rev. Mol. Cell Biol.*, vol. 4, no. 3, pp. 243 –249, 2003. [Online]. Available: http://www.basis.ncl.ac.uk/

[25] M. L. Blinov, J. R. Faeder, B. Goldstein, and W. S. Hlavacek, "BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains," *Bioinformatics*, vol. 20, no. 17, pp. 3289–3291, 2004. [Online]. Available: http://cellsignaling.lanl.gov/bionetgen/

[26] P. Dhar, C. M. Tan, S. Somani, L. Ye, A. Sairam, M. Chitre, Z. Hao, and K. Sakharkar, "Cellware—a multi-algorithmic software for computational systems biology," *Bioinformatics*, vol. 20, no. 8, pp. 1319–1321, 2004. [Online]. Available: http://www.bii.a-star.edu.sg/sbg/cellware/

[27] M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C. Venter, and C. A. H. III, "E-CELL: software environment for whole-cell simulation," *Bioinformatics*, vol. 15, no. 1, pp. 72–84, 1999. [Online]. Available: http://ecell.sourceforge.net/

[28] [Online]. Available: http://www.mathworks.com/products/simbiology/

[29] M. Ander, P. Beltrao, D. V. B., J. Ferkinghoff-Borg, M. Foglierini, A. Kaplan, C. Lemerle, I. Tomás-Oliveira, and L. Serrano, "SmartCell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks," *Syst. Biol.*, vol. 1, pp. 129–138, 2004. [Online]. Available: http://smartcell.embl.de/

[30] A. M. Kierzek, "STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm," *Bioinformatics*, vol. 18, no. 3, pp. 470–481, 2002. [Online]. Available: http://www.sysbio.pl/stocks/

[31] P. Mendes and D. B. Kell, "Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation," *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.