bioPN: An R library of C functions for simulation of stochastic biochemical reactions using Petri Nets

Roberto L. Bertolusso[1], and Marek Kimmel[1,2]

***Short Abstract* — bioPN is a R library of C functions that can be used to simulate time-dependent evolution of biochemical reactions. The model is defined as a place/transition Petri Net. It can either be deterministically solved using an explicit Runge Kutta Dormand Prince 45 method, or simulated using an optimized Gillespie algorithm, or executed as a hybrid of the two, according to the Haseltine and Rawlings' algorithm. The user indicates which reactions are considered slow (stochastic) and fast (deterministic). The library has been designed for speed and flexibility.**

Analytical solutions to most but the basic stochastic systems of biochemical reactions in cells are intractable. This makes necessary to use stochastic simulations. For better approximations, high numbers of runs are needed, so that optimized methods are necessary to perform the task in an acceptable time. In the case of eukaryotic cells the number of reactions of protein production and degradation is so high that exact methods, such as the SSA [1] may become impractical. Several approximations have been developed to speed-up algorithms (for review see Chapter 8 of Wilkinson (2006) [2]). A non-exhaustive review of the available software follows. STOCKS [3] is a non-optimized C++ implementation of the Gillespie SSA method. COPASI [4] is a multi-platform C++ application that allows pure stochastic and pure deterministic simulation. The software also imports SBML files. StochKit [5] is a C++ application that includes the SSA algorithm and approximate stochastic methods. The model is defined in a SBML file. GillespieSSA [6] is an R package programmed in R capable of performing direct and approximate stochastic methods. The user defines structures for the constants and the stoichiometry matrix and provides expressions for reaction propensities. As a means of representing the biochemical networks, we chose the paradigm of Petri Nets (reviewed in Chapter 2 of Wilkinson (2006) [2]) as it is closely related to the way reactions are defined. R [7] was chosen as the backbone because it is a strong language for statistical data analysis that allows post-processing of the results and has plotting capabilities. It is licenced under the GPL, available free of cost, runs natively on several architectures (including Unix, Linux, Windows, and Mac), and has an active and growing developer and user community. It is a slow language because it is interpreted. However, this limitation does not apply in our case, as the bioPN functions run entirely in a compiled C code. Moreover, R provides a C library that can be called from compiled code. Uniform random numbers are generated by the R library functions unif_rand() and rexp(), which have passed strict tests for randomness. BioPN allows the user to choose a specific generator, or use R's default, currently the Mersenne-Twister [8]. In the deterministic and hybrid mode [9] fast reactions are approximated with an explicit Runge Kutta Dormand Prince 45 method [10].

The package has been successfully used in large-scale systems. We present as an example the cross talk between regulatory modules of p53 and NF-kB [11]. This combined model has 46 places and 66 transitions (15 of them slow). One run performed using an original Matlab take about 15 minutes on high end hardware. The corresponding run using bioPN takes about 1.6 seconds on a laptop, making possible 10,000 runs in about 4.6 hours (instead of 104 days). A purely stochastic run takes about 40 seconds, at the rate of about 8 million reactions per second.

[1]Department of Statistics, Rice University, 6100 Main Street, MS138, Houston, TX 77005, USA, E-mail: roberto.bertolusso@rice.edu

[2]Systems Engineering Group, Silesian University of Technology, 44-100 Gliwice, Poland

References

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

2) Wilkinson, D. (2006). Stochastic modeling for systems biology. *Chapman \& Hall/CRC*

3) Kierzek, A. (2002) STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics*, **18**, 470-481.

4) Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P., and Kummer, U. (2006). COPASI. a COmplex PAthway Simulator. *Bioinformatics* **22**, 3067-74

5) Li, H., Cao, Y., Petzold, L, and Gillespie, D. (2007) Algorithms and Software for Stochastic Simulation of Biochemical Reacting Systems. *Biotechnol Prog.* **24**, 56-61.

6) Pineda-Krch, M. (2008) GillespieSSA: Implementing the Gillespie Stochastic Simulation Algorithm in R. *Journal of Statistical Software* **25**, 1-18.

7) R Development Core Team (2008) R: A Language and Environment for Statistical Computing. *ISBN 3-900051-07-0*.

8) Matsumoto, M, and Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, *ACM Trans. on Modeling and Computer Simulation* **8**, 3-30.

9) Haseltine, E, and Rawlings, J. (2002). Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics* **117**, 6959-6969.

10) Dormand, J, and Prince, P. (1980). A family of embedded Runge-Kutta formulae. *Comput. Appl. Math.***6**, 19-26.

11) Puszynski, K., Bertolusso, R., Lipniacki, T. (2009) Crosstalk between p53 and NF-kB systems: pro- and anti-apoptotic functions of NF-kB, *IET Systems Biology*. Accepted for publication.