BIOINFORMATICS APPLICATIONS NOTE Vol. 17 no. 6 2001
Pages 575–576

StochSim: modelling of stochastic biomolecular processes

Nicolas Le Novère and Thomas Simon Shimizu*

Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Received on November 11, 2000; revised on February 27, 2001; accepted on March 1, 2001

ABSTRACT

Summary: StochSim is a stochastic simulator for chemical reactions. Molecules are represented as individual software objects that react according to probabilities derived from concentrations and rate constants. Version 1.2 of StochSim provides a novel cross-platform graphical interface written in Perl/Tk. A simple two-dimensional spatial structure has also been implemented, in which nearest-neighbour interactions of molecules in a 2-D lattice can be simulated.

Availability: Various ports of the program can be retrieved at ftp://ftp.cds.caltech.edu/pub/dbray/

Contact: nl223@cus.cam.ac.uk; tss26@cam.ac.uk

StochSim ALGORITHM

Most available biochemical simulators are deterministic, i.e. they use a set of differential equations to represent the reactions and solve the system iteratively by numerical integration. Such programs include, for instance, Gepasi (Mendes and Kell, 1998), Jarnac (Sauro, 2000) and DBSolve (Goryanin *et al.*, 1999). However, the assumptions made by the deterministic method do not hold true for many intracellular processes which are sensitive to the behaviour of a relatively small number of molecules. The stochastic behaviour of the component molecules must then be considered.

Stochastic simulation of biochemical reactions was pioneered by Gillespie, who developed an efficient algorithm for this purpose (Gillespie, 1976). But, as with conventional deterministic methods, the Gillespie method treats molecular species in bulk and does not represent individual molecules separately. This can be a problem when a system contains proteins that have multiple modification sites which alter their catalytic activity, binding affinity and so on. Each state may have different reactive properties and a detailed model may contain millions of distinct chemical reactions. Both the deterministic method and the Gillespie method, in which the time required for simulation increases proportionally with the number of reactions, have difficulty in coping with these conditions.

In StochSim, each molecule exists as an independent software object. This allows the representation of molecules that have specific internal states, called *multistate molecules*. These react with different probabilities according to the state they are in, and are used to reflect intramolecular events such as covalent modification or conformational changes. The state of a multistate molecule can change over the course of a simulation as a result of the defined reactions. This individual representation allows more accurate modelling of biochemical processes at the cost of higher memory and CPU consumption. See Morton-Firth and Bray (1998) and StochSim manual (Le Novère *et al.*, 2000) for an in-depth discussion of the relative efficiency of the algorithms used by Gillespie and StochSim.

StochSim reduces all possible reactions to elementary unimolecular and bimolecular reactions. The probabilities of all reactions are precomputed based on user-specified rate constants, and stored in a look-up table during initialisation. The length of time that each simulation iteration represents, as well as the probability of unimolecular and bimolecular reactions occuring within each iteration, are also adjusted during initialisation to ensure correct reaction rates. Once these have been set up, the simulation proceeds by iteration of a very simple subroutine in which either a unimolecular or a bimolecular reaction can occur between randomly chosen molecules.

TkStochSim

The original StochSim simulator was written for Microsoft Windows and was provided with a Graphical User Interface (GUI) based on the Microsoft MFC widgets. These widgets were not compatible with other operating systems such as UNIX, Linux, or MacOS. Therefore to run StochSim under such operating systems, the user was required to write the simulation initialisation files manually. This is tedious and error-prone. In order to ease the creation and modification of simulation configura-

*To whom correspondence should be addressed.

tions under a variety of operating environments, a novel interface written in Perl using the Tk widgets has been developed. Not all of the functionalities of STOCHSIM are currently covered. In particular, the configuration of dynamic values have yet to be implemented. The configuration of models using the novel spatial extensions to STOCHSIM (see next paragraph) are also not supported yet. To access these functions of STOCHSIM, the user can directly write the configuration files. However, once the configuration files are written, TkSTOCHSIM can read and run any simulation and display the results.

SPATIAL EXTENSIONS TO STOCHSIM

STOCHSIM 1.0 treated the entire reaction system as a uniformly mixed solution. The omission of spatial heterogeneity has been a norm in biochemical simulations because it greatly facilitates modelling and reduces the computational load of simulation. However, as the resolution of our understanding of biochemical processes increases, it is becoming clear that the spatial organisation of molecules often play an important role. We have therefore undertaken extending STOCHSIM to incorporate a spatial representation.

In the current version (STOCHSIM 1.2), a simple two-dimensional spatial representation has been implemented, in which multistate molecules are treated as probabilistic cellular automata. Nearest-neighbour interactions between molecules in a lattice (such as clustered receptors on a membrane) can be defined to simulate the effect of spatial interactions. This was motivated by studies of the bacterial chemotaxis receptor complex which suggested that lateral interactions between neighbouring complexes exist (Shimizu *et al.*, 2000). Similar clustering behaviour has also been observed for other receptors (Yin and Lai, 2000).

In STOCHSIM 1.2's 2-D complex array, reactions that are sensitive to the state of neighbouring complexes can be defined. The rates of these reactions depend not only on the state of the complex chosen in the STOCHSIM iteration, but also on the states of its nearest neighbours within the complex array. Graphical snapshots of the state of the complex array, as well as traces of the state of individual complexes within the array can be outputted at user-specified intervals.

IMPLEMENTATION

The STOCHSIM simulation engine is written in C^{++}, and has been successfully compiled under Windows95/98 as well as Linux (2.2) running on Alpha and Intel (x86) chips.

The Perl/Tk interface requires at least Perl 5.004 and Perl/Tk 800.022. It has been tested under Windows 98 and Debian GNU/Linux *Slink*, *Potato* and *Woody*.

REFERENCES

Gillespie,D.T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, **22**, 403–434.

Goryanin,I., Hodgman,T.C. and Selkov,E. (1999) Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics*, **15**, 749–758.

Le Novère,N., Shimizu,T.S., Firth,C. and Bray,D. (2000) STOCHSIM, the stochastic simulator. ftp://ftp.cds.caltech.edu/pub/dbray/.

Mendes,P. and Kell,D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, **14**, 869–883.

Morton-Firth,C.J. and Bray,D. (1998) Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.*, **192**, 117–128.

Sauro,H. (2000) Jarnac: an interactive metabolic systems language. http://www.members.tripod.co.uk/sauro/Jarnac.htm.

Shimizu,T.S., Le Novère,N., Levin,M.D., Beavil,A.J., Sutton,B.J. and Bray,D. (2000) Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. *Nat. Cell Biol.*, **2**, 792–796.

Yin,C.C. and Lai,A.F. (2000) Intrinsic lattice formation by the ryanodine receptor calcium-release channel. *Nat. Cell Biol.*, **2**, 669–671.