

Systems biology

# Stochastic reaction-diffusion simulation with MesoRD

Johan Hattne<sup>†</sup>, David Fange and Johan Elf<sup>\*</sup>

Department of Cell and Molecular Biology, BMC, Uppsala University, 75124 Uppsala, Sweden

Received on February 10, 2005; revised and accepted on April 4, 2005

Advance Access publication April 7, 2005

## ABSTRACT

**Summary:** MesoRD is a tool for stochastic simulation of chemical reactions and diffusion. In particular, it is an implementation of the next subvolume method, which is an exact method to simulate the Markov process corresponding to the reaction-diffusion master equation.

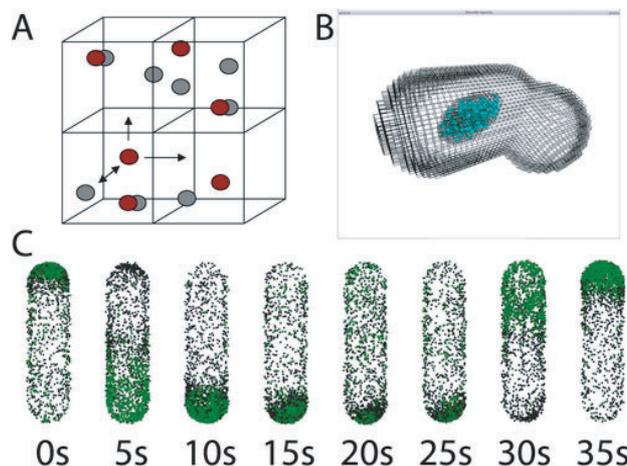
**Availability:** MesoRD is free software, written in C++ and licensed under the GNU general public license (GPL). MesoRD runs on Linux, Mac OS X, NetBSD, Solaris and Windows XP. It can be downloaded from <http://mesord.sourceforge.net>.

**Contact:** johan.elf@icm.uu.se; johan.hattne@embl-hamburg.de

**Supplementary information:** 'MesoRD User's Guide' and other documents are available at <http://mesord.sourceforge.net>.

## 1 INTRODUCTION

Slow intracellular diffusion rates, in combination with reaction rates that depend non-linearly on local molecule concentrations, make stochastic reaction-diffusion descriptions necessary for quantitative analysis of *in vivo* biochemistry (Elf and Ehrenberg, 2004; Lipkow *et al.*, 2005; Schnell and Turner, 2004). A starting point for such a description is the reaction-diffusion master equation (RDME) (Baras and Mansour, 1997; Kuramoto, 1974). In the RDME the total system volume is divided into a number of virtual subvolumes that are small enough to be considered homogenized by diffusion on the time-scale of the chemical reaction. The state of the system is defined by the copy number of each species in each subvolume. This state is changed by chemical reactions within the subvolumes or diffusion jumps of single molecules between neighbouring subvolumes (Fig. 1). The rates of the chemical reactions depend on the local concentrations of the molecules within each subvolume. The rates of diffusion events are modelled as first order reactions, defined in the subvolume where the jump originates. Given the state description and the rates for the different events, the RDME describes the time evolution of the probability of each possible state of the system. Due to the exceptionally high dimension of the state space, neither analytical nor direct numerical solutions of the RDME are possible for interesting biochemical systems. The natural escape route is Monte Carlo simulation of individual realizations of the Markov processes described by the RDME. This kind of simulation was pioneered in 1979 (Malek-Mansour and Houard, 1979), when Gillespie's direct method (Gillespie, 1976) was used to simulate a system with 21 subvolumes in one dimension. Now, however, we are interested in chemical reactions in structured geometries in three dimensions, which may require millions of subvolumes to be correctly discretized. Specialized simulation



**Fig. 1.** (A) A very small example of spatial discretization in  $2 \times 2 \times 1$  subvolumes. (B) Viewer screenshot. (C) Min system oscillations in *E.coli* (Huang *et al.*, 2003) simulated using MesoRD (manuscript in preparation).

methods must be used to obtain reasonable execution times for these systems.

One such method, the next subvolume method (NSM) (Elf and Ehrenberg, 2004), was recently developed. The algorithm behind NSM takes advantage of the special structure of the RDME, and generates trajectories that are equivalent to those from Gillespie's method. Efficient and flexible implementation of the NSM is finicky, which is why we now present MesoRD (released on September 3, 2004).

## 2 MesoRD

### 2.1 Running MesoRD

The kinetic model is read into the MesoRD program at runtime. The model is defined in the systems biology markup language (SBML) (Hucka *et al.*, 2003), with geometry and diffusion information added in annotation fields. Parameters that do not describe the chemical system but still are needed for its simulation, such as spatial discretization, simulation duration and visualization options, are specified at runtime through the user interfaces. Currently, the Windows version only has a graphical user interface, while Unix versions rely solely on the command line for user interaction. Command line control allows for scripted batch processing.

A simulation starts once the SBML file has been parsed, the neighbour relationships among the subvolumes have been defined and the initial number of molecules have been distributed. While simulating,

<sup>\*</sup>To whom correspondence should be addressed.

<sup>†</sup>Present address: EMBL, Notkestrasse 85, 22603 Hamburg, Germany.

the program divides the workload into three threads: (i) The simulation thread, which samples the time to the next reaction or diffusion event from the state dependent probability density functions and updates the state of the system accordingly. (ii) The optional visualization thread, which runs a three-dimensional OpenGL viewer of the simulation. (iii) The status thread, which displays the progress to the user and handles requests to terminate simulation. At user-specified intervals, the full state of the system is written to file for statistical post-processing.

## 2.2 Geometry in MesoRD

In MesoRD, the geometry of each compartment is described using constructive solid geometry (CSG) (Requicha and Tivole, 1978) (<http://hdl.handle.net/1802/1209>). This means that each compartment is defined by differences, intersections or unions of a number of geometric primitives (boxes, cones, cylinders and spheres). Any primitive, or combination of primitives, can be rotated, scaled and translated. Periodic boundary conditions can be used for boxes and cylinders.

## 2.3 Expression evaluation

Rate expressions are specified for each reaction in the SBML model definition. MesoRD handles units explicitly and converts the rates into number of events per time and subvolume. This is required to interpret  $dt \times \text{rate}$  as the probability that the event occurs during an infinitesimal time  $dt$ . MesoRD has a few parameters which take their values at runtime but still can be used in rate expression definitions. One such example is the concentration of one molecule in one subvolume,  $\Omega^{-1}$ , which is necessary to write the rate expression for the reaction  $A + A \xrightarrow{k} \emptyset$   $k \cdot a \cdot (a - \Omega^{-1})$ .

MesoRD represents general mathematical expressions in abstract syntax trees (ASTs). Recursive traversal of such trees can easily turn into a performance bottleneck. MesoRD attempts to restructure ASTs into sums of products that can be evaluated iteratively. For instance, rate expressions derived from the law of mass action can be simplified in this way.

Furthermore, the outcome of any evaluation is cached, so that a subsequent computation of the same rate expression for the same configuration of reactant molecules need not involve any evaluation at all. The cache should yield an acceptable hit rate when the number of reactant molecule configurations is reasonably low. In many cases this is expected to be true; if the number of subvolumes is larger than the number of molecules, the number of molecules of each species within a subvolume is low.

## 2.4 Performance issues

No thorough performance analysis of MesoRD has been made. However, on a Xeon 3.06 GHz with 1 GB RAM, MesoRD simulates approximately  $5 \times 10^8$  reaction or diffusion events per hour for a system with  $10^6$  subvolumes. With five species and five different reactions, MesoRD uses  $\sim 250$  MB for a  $10^6$  subvolume simulation. On dual processor machines with shared memory, visualization does not significantly degrade simulation performance.

The NSM (Elf and Ehrenberg, 2004) scales logarithmically with the number of subvolumes, whereas an implementation of Gillespie's direct method would scale linearly. An implementation of the next reaction method (NRM) (Gibson and Bruck, 2000) would also scale logarithmically with the number of subvolumes. However, the memory requirements and number of operations per iteration would be much higher. The SmartCell software (Ander et al., 2004), which in some aspects is similar to MesoRD, therefore recently changed its simulation algorithm from NRM to NSM.

## 3 CONCLUSIONS

The use of MesoRD is not limited to the realm of molecular biology. Its efficiency for systems with millions of subvolumes and molecules makes it generally useful in physical chemistry. MesoRD can also be used for testing approximations of the RDME based on renormalization group approaches (Lee and Cardy, 1995) and for comparison of RDME descriptions with simulations based on discretization in time instead of space (Andrews and Bray, 2004).

## ACKNOWLEDGEMENTS

MesoRD is dependent on a number of Open Source community efforts that are acknowledged in the MesoRD documentation. The development of MesoRD was funded by grants to Måns Ehrenberg from the Swedish Research Council.

## REFERENCES

- Ander, M. et al. (2004) SmartCell: a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple gene networks. *Syst. Biol.*, **1**, 129–138.
- Andrews, S.S. and Bray, D. (2004) Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.*, **1**, 137–151.
- Baras, F. and Mansour, M.M. (1997) Microscopic simulation of chemical instabilities. *Adv. Chem. Phys.*, **100**, 393–475.
- Elf, J. and Ehrenberg, M. (2004) Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Syst. Biol.*, **2**, 230–236.
- Gibson, M. and Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many species and channels. *J. Phys. Chem. A*, **104**, 1876–1889.
- Gillespie, D. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, **22**, 403–434.
- Huang, K.C. et al. (2003) Dynamic structures in *Escherichia coli*: spontaneous formation of MinE rings and MinD polar zones. *Proc. Natl Acad. Sci. USA*, **100**, 12724–12728.
- Hucka, M. et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, **19**, 524–531.
- Kuramoto, Y. (1974) Effects of diffusion on the fluctuations in open systems. *Prog. Theor. Phys.*, **52**, 711–713.
- Lee, B. and Cardy, J. (1995) Renormalization-group study of the  $A + B \rightarrow \phi$  diffusion-limited reaction. *J. Stat. Phys.*, **80**, 971–1007.
- Lipkow, K. et al. (2005) Simulated diffusion of phosphorylated CheY through the cytoplasm of *Escherichia coli*. *J. Bacteriol.* **187**, 45–53.
- Malek-Mansour, M. and Houard, J. (1979) A new approximation scheme for the study of fluctuations in nonuniform nonequilibrium systems. *Phys. Lett. A*, **70**, 366–368.
- Requicha, A.A.G. and Tivole, R.B. (1978) Mathematical foundations of constructive solid geometry. <http://hdl.handle.net/1802/1209>
- Schnell, S. and Turner, T.E. (2004) Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. Biophys. Mol. Biol.*, **85**, 235–260.