

SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations

Jason Lloyd-Price*, Abhishekh Gupta and Andre S. Ribeiro

Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland

Associate Editor: Martin Bishop

ABSTRACT

Motivation: Cell growth and division affect the kinetics of internal cellular processes and the phenotype diversity of cell populations. Since the effects are complex, e.g. different cellular components are partitioned differently in cell division, to account for them in silico, one needs to simulate these processes in great detail.

Results: We present SGNS2, a simulator of chemical reaction systems according to the Stochastic Simulation Algorithm with multi-delayed reactions within hierarchical, interlinked compartments which can be created, destroyed and divided at runtime. In division, molecules are randomly segregated into the daughter cells following a specified distribution corresponding to one of several partitioning schemes, applicable on a per-molecule-type basis. We exemplify its use with six models including a stochastic model of the disposal mechanism of unwanted protein aggregates in *Escherichia coli*, a model of phenotypic diversity in populations with different levels of synchrony, a model of a bacteriophage's infection of a cell population and a model of prokaryotic gene expression at the nucleotide and codon levels.

Availability: SGNS2, instructions and examples available at www.cs.tut.fi/~lloydprice/sgns2/ (open source under New BSD license).

Contact: jason.lloyd-price@tut.fi

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 31, 2012; revised on August 30, 2012; accepted on September 6, 2012

1 INTRODUCTION

Recent evidence suggests that even in cellular organisms whose division is morphologically symmetric, there are a number of asymmetries between daughter cells. These arise, among other things, from the stochasticity in the partitioning of components in division (Huh and Paulsson, 2011) and from biased partitioning schemes for some components. For example, in *Escherichia coli*, unwanted protein aggregates follow biased partitioning schemes dependent on the age of the daughter cells' poles (Lindner *et al.*, 2008).

These and other recent findings suggest that the phenotypic diversity of cell populations, among other factors, depends on errors and biases in the partitioning of RNA, proteins and other molecules. This is of relevance since most RNAs exist in small numbers (Bernstein *et al.*, 2002) and small fluctuations in these numbers can alter the behavior of genetic circuits (Ribeiro and Kauffman, 2007) and trigger visible phenotype changes (Choi *et al.*, 2008).

These sources of phenotypic heterogeneity are difficult to distinguish from, e.g. noise in gene expression (Huh and Paulsson, 2011). Although some effects can be assessed analytically (Huh and Paulsson, 2011), others are too complex and must be assessed numerically. A simulator is thus needed that accounts for noise and delays (Kandhavelu *et al.*, 2012) in gene expression and for compartmentalization of processes and components.

Presently, simulators of the dynamics of noisy biochemical systems rely on the Stochastic Simulation Algorithm (SSA) (Gillespie, 1977), e.g. (Blakes *et al.*, 2011; Hattne *et al.*, 2005; Hoops *et al.*, 2006; Lok and Brent, 2005). Some support compartmentalization, simulating reaction-diffusion systems in either static (Hattne *et al.*, 2005) or dynamically sized compartments (Blakes *et al.*, 2011; Versari and Busi, 2008). Others support rule-based creation of reactions at runtime (Lok and Brent, 2005; Spicher *et al.*, 2008), and thus can simulate a dynamic cell population. Very few support delays on the release into the system of one or more products of a reaction (Roussel and Zhu, 2006). These delays are essential to accurately model the kinetics of some processes, e.g. transcription, as RNA production is mostly regulated by the duration of events in transcription initiation (Muthukrishnan *et al.*, 2012).

Here, we present SGNS2, an extension of SGN Sim (Ribeiro and Lloyd-Price, 2007) that incorporates dynamic compartments and multiple partitioning distributions at cell division, applicable on a per-molecule-type basis.

2 METHODS

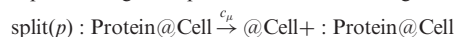
SGNS2 is an extension of SGNS, the stochastic simulator of SGN Sim (Ribeiro and Lloyd-Price, 2007). It contains all the features of SGNS, such as reactions with multi-delayed events. The two key additions in SGNS2 are (i) it supports dynamic, interlinked, hierarchical compartments and (ii) it supports multiple molecule and compartment partitioning schemes, applicable on a per-molecule-type basis. The novel features considerably extend the class of models that can be simulated.

SGNS2 uses a modified version of the Next Reaction Method (NRM) (Gibson and Bruck, 2000). Namely, the NRM was adapted to stochastic P-systems (Spicher *et al.*, 2008) by using a hierarchy of indexed priority queues (IPQ, an ordered list of elements that keep track of their position in the list) and further modified to allow multiple delays in reactions. The IPQ data structure, implemented with a binary heap, is described in Gibson and Bruck (2000). We use a separate IPQ for each compartment, which publish a 'tentative next event time' to an overall IPQ which determines the next event time in the entire simulation. We optimize the update step when molecule populations in a parent compartment change by using a hierarchical refinement of the IPQs with appropriate scaling of tentative firing times (see Supplementary Material). Delayed events were implemented by creating wait lists, implemented by binary heap-based priority queues, whose earliest event is published to each compartment's indexed priority queue. The simulation's elementary SSA steps scale logarithmically with

*To whom correspondence should be addressed.

the number of reactions, compartments and delayed events, allowing complex models to be simulated in reasonable time.

To simulate cell division, we introduced a special reaction event, whose timing follows the SSA rules. When executed, instead of subtracting substrates from the system, a random number is generated based on one of the several partitioning distributions available, including some of those listed in Huh and Paulsson (2011). Each of these mimics a specific molecule partitioning process during cell division. SGNS2 allows both biased and unbiased partitioning of molecules and sub-compartments. The results of these events can be instantaneous or be placed on the wait list. Compartment division and molecule partitioning are represented in the following form:



When this reaction occurs, a new cell compartment is created (@Cell in the product list). Proteins in the original cell are partitioned according to a biased binomial partitioning scheme. In this, each protein is independently partitioned into the new cell with probability p . Other common partitioning distributions include the independent partitioning of molecules into daughter cells with random (beta-distributed) sizes and the binding of molecules to spindle binding sites which are segregated evenly between daughter cells such as during mitosis. Available distributions are listed in the manual.

SGNS2 is a command line utility, designed to fit into a toolchain, supporting various input and output formats. Input can be specified in two formats: SBML (Hucka *et al.*, 2003) and SGNSim's native format (Ribeiro and Lloyd-Price, 2007). A subset of SBML Core level 3 version 1 is supported, allowing simulation of most SBML models. Output can be in csv, tsv or in binary format. A text editor may be used to write models in SGNSim format. SBML-based graphical interfaces such as CellDesigner (Funahashi *et al.*, 2008) or Cytoscape (Smoot *et al.*, 2011) may be used to manage SBML models. The results of simulations are interpretable by programs like MATLAB, R or Excel. An example of running a model in SGNSim format of a growing cell population is shown in Supplementary Figure S1.

3 DISCUSSION

SGNS2 is the first stochastic simulator that includes multi-delayed events, dynamic compartments and molecule partitioning schemes in division. To test its correctness, we simulated models from the Discrete Stochastic Model Test Suite (Evans *et al.*, 2008). All showed the expected behavior (Supplementary Figs S2 and S3).

SGNS2, though making use of existing and slightly modified versions of existing algorithms, can simulate an array of biological processes not previously possible. For example, it is ideal for simulating gene expression at the nucleotide and codon levels (see 'Availability' section) and study features such as how events in transcription elongation affect protein production kinetics (Mäkelä *et al.*, 2011).

SGNS2 is also suited to study partitioning in cell division, which affects aging, among other processes, and is of particular relevance when modeling populations over multiple generations. To exemplify this, we modeled the biased partitioning of protein aggregates in *E. coli*, known to accumulate in cells with older poles, reducing vitality (Lindner *et al.*, 2008). The results in Supplementary Figure S4 agree with measurements (Stewart *et al.*, 2005). We further studied how cell-cycle synchrony affects the population-level statistics of RNA numbers [Supplementary Fig. S5, in agreement with measurements in Lloyd-Price *et al.* (2012)]. As a side note, we expect the partitioning of RNA and proteins to affect the dynamics of genetic circuits, particularly the stability of their noisy attractors (Ribeiro and Kauffman, 2007). To further demonstrate the simulator's utility, we modeled the viral infection of a dynamic bacterial population.

In conclusion, SGNS2 provides novel functionalities to model and simulate cellular processes not previously possible, as seen from the examples. In general, SGNS2 enables the modeling of stochastic processes in live cells that require compartmentalization, multi-delayed complex processes and complex stochastic partitioning schemes at a per-molecule type in cell division. These features are necessary to study *in silico*, among other phenomena, phenotypic diversity in cell populations.

Funding: Work supported by Academy of Finland (126803).

Conflict of Interest: none declared.

REFERENCES

- Bernstein, J.A. *et al.* (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.
- Blakes, J. *et al.* (2011) The infobiotics workbench: an integrated *in silico* modelling platform for systems and synthetic biology. *Bioinformatics*, **27**, 3323–3324.
- Choi, P.J. *et al.* (2008) A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.
- Evans, T. *et al.* (2008) The SBML discrete stochastic models test suite. *Bioinformatics*, **25**, 285–286.
- Funahashi, A. *et al.* (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE*, **96**, 1254–1265.
- Gibson, M.A. and Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A*, **104**, 1876–1889.
- Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Hattne, J. *et al.* (2005) Stochastic reaction-diffusion simulation with mesord. *Bioinformatics*, **21**, 2923–2924.
- Hoops, S. *et al.* (2006) Copasi a complex pathway simulator. *Bioinformatics*, **22**, 3067–3074.
- 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.
- Huh, D. and Paulsson, J. (2011) Random partitioning of molecules at cell division. *Proc. Natl Acad. Sci. USA*, **108**, 15004–15009.
- Kandhavelu, M. *et al.* (2012) Random partitioning of molecules at cell division. *Phys. Biol.*, **9**, 026004.
- Lindner, A.B. *et al.* (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl Acad. Sci. USA*, **105**, 3076–3081.
- Lloyd-Price, J. *et al.* (2012) Probabilistic RNA partitioning generates transient increases in the normalized variance of RNA numbers in synchronized populations of *Escherichia coli*. *Mol. Biosys.*, **8**, 565–571.
- Lok, L. and Brent, R. (2005) Automatic generation of cellular reaction networks with molecularizer 1.0. *Nat. Biotech.*, **23**, 131–36.
- Mäkelä, J. *et al.* (2011) Stochastic sequence-level model of coupled transcription and translation in prokaryotes. *BMC Bioinf.*, **12**, 121.
- Muthukrishnan, A.-B. *et al.* (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res.*, **40**, 8472–8483.
- Ribeiro, A.S. and Kauffman, S. (2007) Noisy attractors and ergodic sets in models of gene regulatory networks. *J. Theor. Biol.*, **247**, 743–755.
- Ribeiro, A.S. and Lloyd-Price, J. (2007) SGN sim, a stochastic genetic networks simulator. *Bioinformatics*, **23**, 777–779.
- Roussel, M.R. and Zhu, R. (2006) Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys. Biol.*, **3**, 274–284.
- Smoot, M. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
- Spicher, A. *et al.* (2008) Stochastic P systems and the simulation of biochemical processes with dynamic compartments. *Biosystems*, **91**, 458–472.
- Stewart, E.J. *et al.* (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, **3**, e45.
- Versari, C. and Busi, N. (2008) Efficient stochastic simulation of biological systems with multiple variable volumes. *Elec. Notes Theor. Comp. Sci.*, **194**, 165–180.