BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 24 2007, pages 3409–3411
doi:10.1093/bioinformatics/btm491

Systems biology

CellLine, a stochastic cell lineage simulator

Andre S. Ribeiro*, Daniel A. Charlebois and Jason Lloyd-Price

Institute for Biocomplexity and Informatics, University of Calgary, Canada

Received on July 23, 2007; revised on August 20, 2007; accepted on September 23, 2007

Advance Access publication October 9, 2007

Associate Editor: Limsoon Wong

ABSTRACT

Summary: We present *CellLine*, a simulator of the dynamics of gene regulatory networks (GRN) in the cells of a lineage. From user-defined reactions and initial substance quantities, it generates cell lineages, i.e. genealogic pedigrees of cells related through mitotic division. Each cell's dynamics is driven by a delayed stochastic simulation algorithm (delayed SSA), allowing multiple time delayed reactions.

The cells of the lineage can be individually subject to 'perturbations', such as gene deletion, duplication and mutation. External interventions, such as adding or removing a substance at a given moment, can be specified. Cell differentiation lineages, where differentiation is stochastically driven or externally induced, can be modeled as well. Finally, *CellLine* can generate and simulate the dynamics of multiple copies of any given cell of the lineage.

As examples of *CellLine* use, we simulate the following systems: cell lineages containing a model of the P53-Mdm2 feedback loop, a differentiation lineage where each cell contains a 4 gene repressilator (a bistable circuit), a model of the differentiation of the cells of the retinal mosaic required for color vision in *Drosophila melanogaster*, where the differentiation pathway depends on one substance's concentration that is controlled by a stochastic process, and a 9 gene GRN to illustrate the advantage of using *CellLine* rather than simulating multiple independent cells, in cases where the cells of the lineage are dynamically correlated.

Availability: The *CellLine* program, instructions and examples are available at www.ucalgary.ca/~aribeiro/CellLine/CellLine.html

Contact: ARibeiro@ucalgary.ca

1 INTRODUCTION

Biological systems can be modeled at many scales: from detailed descriptions of underlying molecular interactions, to abstract models of large scale systems.

It has been shown that stochastic fluctuations play a key role in biological processes, both at the molecular level, e.g. in gene expression dynamics (Elowitz *et al.*, 2002), and at the macroscopic level, e.g. acting as the driving force of cells' choice for phenotypic state (Arkin *et al.*, 2006; Suel *et al.*, 2006).

An example of how the stochastic nature of the underlying chemical processes of gene expression can affect cell differentiation (Arkin *et al.*, 2006) was given (Suel *et al.*, 2006). Using an excitable stochastic model, this work revealed the mechanisms controlling *Bacillus subtilis*' changes from competence to non-competence and vice versa. The results, supported by experiments, show how stochastic fluctuations in GRNs dynamics allow excitable transient changes in a cell's phenotype. Another study (Wernet *et al.*, 2006) showed that the retinal mosaic required for color vision in *Drosophila melanogaster* results from the stochastic expression of a single transcription factor (SS), which determines the differentiation pathway of some of the cells of this mosaic.

Other studies showed that time delays, namely for transcription and translation, should be included (Gaffney and Monk 2006; Zhu *et al.*, 2007) when modeling GRNs. Recently, an extension of the original SSA (Gillespie, 1976) was proposed: the delayed SSA (Roussel and Zhu, 2006). This new algorithm accounts for the stochastic nature of chemical reactions, but can also model multi-step processes in a single step by accounting for the non-negligible time duration for these events to be completed. The use of multiple time delayed reactions to model, e.g. transcription and translation, were proven crucial to correctly mimicking experimental observations of the dynamics of single gene expression at the single protein level (Zhu *et al.*, 2007).

We present *CellLine*, a simulator that can model a cell's lineage from a single cell with as many generations as desired, simulating each cell dynamics during a user-defined lifetime.

Each 'cell' consists of a user-defined set of chemical reactions (e.g. modeling a GRN) and initial substance quantities. The dynamics of each cell is modeled at a detailed level, and therefore, driven by the delayed SSA. *CellLine* can also create and simulate the dynamics of multiple copies of any single cell of the lineage. This is useful to obtain the average dynamics of a population.

Rate constants, time delays or initial concentrations can be defined as variables and their value randomly drawn from a user-defined distribution. Cells can also be individually subject to external perturbations at any desired time during a cell lineage simulation, e.g. the addition or removal of a substance.

Finally, we note that during a cell lineage simulation, in addition to transmitting all substance concentrations at the moment of cell division to the daughter cells, *CellLine* will also transmit any change that occurred in the mother cell's set of reactions during its lifetime (e.g. gene duplication or deletion events). This feature is important in many cases. For example, it allowed us to mimic recent experimental observations that the oscillations in concentrations of P53 and Mdm2, if occurring in the mother cell at the moment of division, are passed on to the daughter cells, and that the oscillations in these two cells remain

*To whom correspondence should be addressed.

synchronized for a significant time interval (Geva-Zatorsky *et al.*, 2006) (see Availability section).

2 METHODS

To simulate the dynamics of the cells of a lineage, first we need to define the set of chemical reactions that can occur within the initial cell and the initial quantities of all the substances in the system. These are placed in a 'reactions file'. Any set of reactions and initial quantities can be defined. As cells divide, the set of reactions in the mother cell is passed on to its daughter cells, as well as the quantities of all the substances at the moment of division.

CellLine consists of two main modules: 'CellLineGen' and 'DynSim'. From the initial reactions file, CellLineGen creates, as cells divide, reactions files for each cell of the lineage. Each cell dynamics is simulated by DynSim. *CellLine's* usage can be described as follows (refer to Availability section for a detailed description): (i) Reactions file: a file containing the set of reactions that can occur within a 'cell' and the initial quantities of the chemical substances in the mother cell. (ii) CellLineGen module: the user sets the number of cell divisions, cell lifetime and sampling interval of the time series of cell's chemical species. CellLineGen calls DynSim to simulate each cell's dynamics as they are generated, and divides each cell into two new cells at the end of a cell lifetime. CellLineGen outputs a reactions file and time series for each cell of the lineage. The total experiment time equals the user defined cell lifetime times the number of generations.

To simulate the dynamics of a particular cell of the lineage multiple times, we provide 'NCellsGen'. This module creates multiple copies of any given single cell, each with a different random seed. Given the number of copies desired, the experiment time duration (i.e. each cell lifetime), and the sampling interval of the chemical species quantities, NCellsGen creates a reactions file for each cell and then calls DynSim to simulate each cell's dynamics and outputs the time series of the chemical substances.

CellLine's dynamics module, DynSim, is driven by the delayed SSA (Roussel and Zhu, 2006). Unlike the original SSA (Gillespie, 1976), the delayed SSA stores delayed output events on a waitlist, and is able to handle multiple delayed output events for each input event. The waitlist is a list of elements, each to be released some time after the reaction that produced them occurred (also stored on the waitlist).

DynSim was built from SGN*Sim* (Ribeiro and Lloyd-Price, 2007), inheriting most of its functionality and sharing the same syntax for writing input files. Delays can be drawn from a distribution at runtime and reaction rates can be a complex function of concentrations at each moment. This is useful to cope with the lack of detailed knowledge of all chemical processes involved in a given system of reactions, thus requiring the use of functions to compute rate constants. Using these functions can also make the simulation faster if it reduces the number of reactions that must occur.

An important feature of *CellLine* is its ability to perturb specific or randomly chosen cells of a lineage, at a given or random moment. The consequences of perturbations are passed on to daughter cells. It is also possible to impose any given initial state on each cell of the lineage. For example, all the proteins present in the cells at the moment of division can be deleted, or the mother cell's initial state can be imposed on daughter cells. Similarly, asymmetric cell division can be modeled.

The output of *CellLine* is a set of tab-delimited text files, one for each cell of the lineage, with the time series of all chemical species of each cell for a user-defined time duration and sampling interval (see Availability section). Fourier spectrums of the time series can also be computed for single cells.

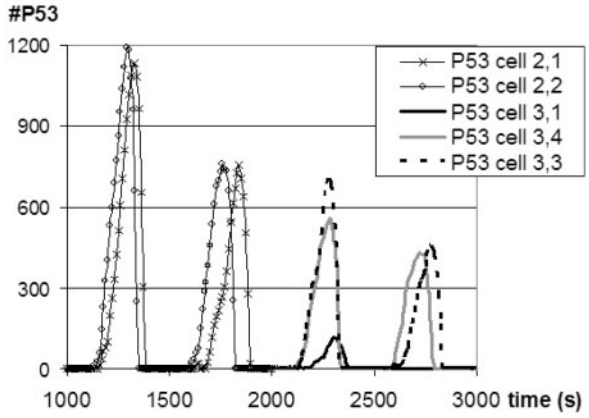

Fig. 1. Time series of P53 in cells of the second and third generation of a lineage. Cell division occurs every 1000 s. Cell (i, j) corresponds to cell number j of generation i. The results match reported measurements (Geva-Zatorsky *et al.*, 2006). Cells (3,3) and (3,4), daughters from the same mother [cell (2,2)], are almost synchronized between themselves but not with cell (3,1), which has a different mother cell [cell (2,1)] (see Availability section).

3 DISCUSSION

CellLine can simulate the dynamics of cells within a lineage, given a reactions file containing a set of chemical reactions and initial substance quantities. As examples of applications, first we simulate a model of the P53-Mdm2 feedback loop that mimics a recent experimental observation that oscillations in the concentrations of proteins P53 and Mdm2, caused by irradiation, are passed on to daughter cells (Geva-Zatorsky *et al.*, 2006). The initial synchronization of the oscillations in the two daughter cells is lost over time. Figure 1 shows one time series of cells from a lineage. *CellLine* also replicated measurements of average P53 concentrations in multiple cells.

Next, we model a 3 gene repressilator and exemplify *CellLine's* ability to impose perturbations at runtime. A 4 gene repressilator, a bistable GRN, is also modeled to demonstrate how to simulate a cell differentiation lineage, where differentiation is driven by stochastic fluctuations in the cell's GRN.

Also, we simulate the dynamics of a model of the cells that determine the color vision mosaic in *D. melanogaster* (Wernet *et al.*, 2006), to show how to use *CellLine* to model a differentiation process where the GRN's final state, and thus the cell's fate, depends on the abundance of a specific chemical substance at the moment of differentiation.

Finally, a cell lineage and a set of multiple independent cells were simulated from an initial cell containing a 9 gene GRN. This is a case where the dynamics of the GRN of cells in a lineage could not be reproduced by simulating multiple independent cells. For simulator, models and results described, see Availability section.

Conflict of Interest: none declared.

REFERENCES

Elowitz,M. *et al.* (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1129–1131.

Gaffney,E.A. and Monk,N.A.M. (2006) Gene expression time delays and turing pattern formation systems. *Bull. Math. Biol.*, **68**, 99–130.

Geva-Zatorsky,N. *et al*. (2006) Oscillations and variability in the p53 system. *Mol. Syst. Biol.*, **2**, doi:10.1038/msb4100068.

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

Ribeiro,A.S. and Lloyd-Price,J. (2007) SGN Sim, a stochastic genetic networks simulator. *Bioinformatics*, **23**, 777–779.

Roussel,M. 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.

Samoilov,M.S. *et al*. (2006) From fluctuations to phenotypes: the physiology of noise. *Sci. STKE*, **366**, re17.

Suel,G. *et al*. (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, **440**, 545–550.

Wernet,M. *et al*. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, **440**, 174–180.

Zhu,R. *et al*. (2007) Studying gene regulatory networks at the molecular level. *J. Theor. Biol.*, **246**, 725–745.