

An Algorithm for the Stochastic Simulation of Gene Expression and Heterogeneous Population Dynamics

Daniel A. Charlebois^{1,2,*}, Jukka Intosalmi^{3,4}, Dawn Fraser^{1,2} and Mads Kærn^{1,2,5*}

¹ Department of Physics, University of Ottawa, 150 Louis Pasteur, Ottawa, Ontario, K1N 6N5, Canada.

² Ottawa Institute of Systems Biology, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada.

³ Department of Mathematics, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland.

⁴ Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland.

⁵ Department of Cellular and Molecular Medicine, University of Ottawa, 451 Smyth Road, Ottawa, Ontario, K1H 8M5, Canada.

Received 28 January 2010; Accepted (in revised version) 7 May 2010

Available online 5 August 2010

Abstract. We present an algorithm for the stochastic simulation of gene expression and heterogeneous population dynamics. The algorithm combines an exact method to simulate molecular-level fluctuations in single cells and a constant-number Monte Carlo method to simulate time-dependent statistical characteristics of growing cell populations. To benchmark performance, we compare simulation results with steady-state and time-dependent analytical solutions for several scenarios, including steady-state and time-dependent gene expression, and the effects on population heterogeneity of cell growth, division, and DNA replication. This comparison demonstrates that the algorithm provides an efficient and accurate approach to simulate how complex biological features influence gene expression. We also use the algorithm to model gene expression dynamics within "bet-hedging" cell populations during their adaptation to environmental stress. These simulations indicate that the algorithm provides a framework suitable for simulating and analyzing realistic models of heterogeneous population dynamics combining molecular-level stochastic reaction kinetics, relevant physiological details and phenotypic variability.

PACS: 87.10.Mn, 87.10.Rt, 87.16.Yc, 87.17.Ee

Key words: Constant-number Monte Carlo, stochastic simulation algorithm, gene expression, heterogeneous population dynamics.

*Corresponding authors. *Email addresses:* daniel.charlebois@uottawa.ca (D. A. Charlebois), mkaern@uottawa.ca (M. Kærn)

1 Introduction

Stochastic mechanisms play key roles in biological systems since the underlying biochemical reactions are subject to molecular-level fluctuations (see, e.g., [11, 28]). Chemical reactions are discrete events occurring between randomly moving molecules. Consequently, the timing of individual reactions is nondeterministic and the evolution of the number of molecules is inherently noisy. One example of particular importance is the stochastic expression of gene products (mRNA and protein) [11, 12, 20, 23, 28]. Here, molecular-level fluctuations may cause genetically identical cells in the same environment to display significant variation in phenotypes, loosely defined as any observable biochemical or physical attribute. While such noise is generally viewed as detrimental due to reduced precision of signal transduction and coordination, several scenarios exist where noise in gene expression may provide a fitness advantage (see Fraser and Kærn [6] for a review). For example, it has been proposed that a cell population may enhance its ability to reproduce (fitness) by allowing stochastic transitions between phenotypes to increase the likelihood that some cells are better positioned to endure unexpected environmental fluctuations [1].

Due to the importance of noise in many biological systems, models involving stochastic formulations of chemical kinetics are increasingly being used to simulate and analyze cellular control systems [9]. In many cases, obtaining analytical solutions for these models is not feasible due to the intractability of the corresponding system of nonlinear equations. Thus, a Monte Carlo (MC) simulation procedure for numerically calculating the time evolution of a spatially homogeneous mixture of molecules is commonly employed [7, 8]. Among these procedures, the Gillespie stochastic simulation algorithm (SSA) is the *de-facto* standard for simulating biochemical systems in situations where a deterministic formulation may be inadequate. The SSA tracks the molecular number of each species in the system as opposed to the variation in concentrations in the deterministic framework. Hence, high network complexity, large separation of time-scales and high molecule numbers can result in computationally intensive executions. Another challenge is the need for simulating cell populations. In many cases, gene expression is measured for 10-100 thousand individuals sampled from an exponentially growing culture of continuously dividing cells. While the dynamics of individual cells can be appropriately simulated by disregarding daughter cells, repeating such simulations for a fixed number of cells may not capture population variability arising from asymmetric division, for example, or age-dependent effects. The alternative, tracking and simulating all cells within the population, is intractable beyond a few divisions due to an exponential increase in CPU demands as a function of time [22].

Here, we present a flexible algorithm to enable simulations of heterogeneous cell population dynamics at single-cell resolution. Deterministic and Langevin approaches to account for changes in intracellular content and the constant-number MC method [18, 31] were previously been combined to simulate and analyze gene expression across cell populations [21, 22]. In these studies, extrinsic heterogeneity associated with stochastic divi-