Computational Biology and Chemistry 56 (2015) 98–108

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem

Research Article

On the impact of discreteness and abstractions on modelling noise in gene regulatory networks

Chiara Bodei[a], Luca Bortolussi[b,g,h], Davide Chiarugi[g,c,*], Maria Luisa Guerriero[d], Alberto Policriti[e], Alessandro Romanel[f]

[a] Dip. di Informatica, Università di Pisa, Italy
[b] Dip. di Matematica e Geoscienze, Università di Trieste, Italy
[c] Max Planck Institut of Colloids and Interfaces, Potsdam, Germany
[d] Systems Biology Ireland, University College Dublin, Ireland
[e] Dip. di Matematica e Informatica, Università di Udine, Udine, Italy
[f] Centre for Integrative Biology (CIBIO), University of Trento, Italy
[g] CNR-ISTI, Pisa, Italy
[h] Modelling and Simulation Group, University of Saarland, Campus E 1 3, Saarbruecken, Germany

ARTICLE INFO

Article history:
Received 13 August 2014
Received in revised form 26 February 2015
Accepted 7 April 2015
Available online 8 April 2015

Keywords:
Gene regulatory networks
Discrete modelling
Hybrid system
Quasi-steady state approximation
Stochastic noise

ABSTRACT

In this paper, we explore the impact of different forms of model abstraction and the role of discreteness on the dynamical behaviour of a simple model of gene regulation where a transcriptional repressor negatively regulates its own expression. We first investigate the relation between a minimal set of parameters and the system dynamics in a purely discrete stochastic framework, with the twofold purpose of providing an intuitive explanation of the different behavioural patterns exhibited and of identifying the main sources of noise. Then, we explore the effect of combining hybrid approaches and quasi-steady state approximations on model behaviour (and simulation time), to understand to what extent dynamics and quantitative features such as noise intensity can be preserved.

1. Introduction

Regulating gene expression is a complex work of orchestration, where the instruments play with improvised variations without a fixed music sheet. Under this regard, the regulation process, in which DNA drives the synthesis of cell products such as RNA, and proteins, can be thought of as a stochastic process. The amount of RNA and proteins in living cells must be thoroughly tuned, both to manage effectively housekeeping functions and to respond promptly to upcoming needs (e.g. to adapt to environmental changes). To this end, gene expression is equipped with several control mechanisms and strategies that grant both reliability and flexibility in terms of throughput. Nevertheless, when observed at the single cell level, the amount of molecules involved in gene expression and its regulation fluctuates randomly (Raj and van Oudenaarden, 2009). This stochastic effect at the molecular level turns out to play important roles in conditioning cell-scale phenomena, e.g. cellular fate decision making, incomplete penetrance or enhanced fitness through phenotypes variability (Raj and van Oudenaarden, 2009).

Pioneering works (Novic and Weiner, 1957; Ross et al., 1994) showed that gene expression in populations of genotypically identical cells (i.e. with the same genetic constitution) is highly variable even when epigenetic conditions (i.e. the ones that result from external rather than genetic influence) are kept constant. In Elowitz et al. (2002), the authors identified such a population variability and decomposed the extrinsic and intrinsic contributions therein. Also, in Murphy et al. (2010), it is shown that this variability it is controllable.

Recent developments in experimental techniques (see Raj and van Oudenaarden (2009) for a review) have made it possible to detect and count individual molecules and, therefore, to measure the amount of mRNA and proteins in single cells. These measurements have clearly shown that the number of mRNA and proteins can vary significantly from cell to cell. This variability is caused

* Corresponding author at: Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, 14476 Potsdam, Germany. Tel.: +49 (331) 567 9617.
E-mail addresses: chiara@di.unipi.it (C. Bodei), luca@dmi.units.it (L. Bortolussi), davide.chiarugi@mpikg.mpg.de (D. Chiarugi), maria.guerriero@astrazeneca.com (M.L. Guerriero), policriti@dimi.uniud.it (A. Policriti), romanel@science.unitn.it (A. Romanel).

http://dx.doi.org/10.1016/j.compbiolchem.2015.04.004
1476-9271/

by the fundamentally stochastic nature of the biochemical events involved in gene expression (Raj and van Oudenaarden, 2009) and is studied, e.g., in Mantzaris (2007), Stamatakis and Zygourakis (2010), Stamatakis and Zygourakis (2011), where population-level mathematical frameworks are introduced and applied.

As a consequence, the phenotypical variability (i.e. the variability resulting from the interaction of the genotype with the environment) exhibited by populations of identical organisms can be directly caused by stochasticity at the single cell level. Thus, it is becoming clear that noise and stochasticity underlie critical events in cell's life such as differentiation and decision making (Balázsi et al., 2011). Moreover, some authors suggest that random phenotypic switching can represent an efficient mechanism for adapting to fluctuating environments (see, e.g., Balázsi et al., 2011; Raj et al., 2010). These findings have raised new interest in analysing the role of noise in gene expression.

The regulation process includes multiple steps leading from gene transcription to the translation of the resulting mRNA to obtain the encoded protein. Each step represents a possible control point, where several biochemical mechanisms play a role (see Alberts et al. (2002) for a comprehensive review and Lillacci and Khammash (2010) for an application to model selection). Characterising the contributions of each single control point in the regulation process of gene expression is a complex task. Identifying which strategies come into play in generating or dampening noisy behaviours is even more challenging. The extensively studied regulation paradigm represented by the feedback control strategy can be used to explain the mechanisms controlling gene transcription and translation. In such mechanisms, the global intensity of the feedback depends on parameters related to every single control point. Computational methods can significantly help investigating the synergistic mechanisms underlying the regulation of gene expression.

There are several modelling strategies that can lead to different kinds of computational models, depending on both the particular purposes and on the features of the available data. Models of biological systems proposed in the literature can vary in terms of the abstraction level used to represent molecular amounts (in this regard a model can be either *discrete* or *continuous*) and in terms of the underlying paradigm used for describing the temporal evolution of the system, which can be either *deterministic* or *stochastic*.

Ordinary Differential Equations (ODEs) have been extensively used over the years to describe the behaviour of biological processes. They provide modellers with powerful and well assessed analysis and simulation techniques. Nevertheless, ODE models implicitly assume continuous and deterministic change of concentrations, abstracting away noise and randomness due to stochastic fluctuations. This, for example, makes it difficult to capture qualitatively different outcomes arising from identical initial conditions (e.g., Raj and van Oudenaarden, 2008; Hume, 2000; Ross et al., 1994).

One way to represent noise is to couple a Gaussian noise term to the model equations, obtaining a set of Stochastic Differential Equations (SDEs). This approach succeeded in gaining insights on the stochasticity of gene expression underlying circadian clocks (Chabot et al., 2007) and genetic switches (Becksei et al., 2001). However, continuous methods still fail to properly describe various phenomena arising from stochastic fluctuations in systems involving small copy numbers of molecules (Resat et al., 2009), as in the case of bimodal mRNA distributions generated by long transcriptional bursts, during which mRNA level approaches a new steady state (Raj and van Oudenaarden, 2009).

The copy numbers of molecules and individual entities in the cell space are discrete and the reactions in which they are involved are stochastic events. Consequently, approaches based on a discrete and stochastic formulation, such as the ones built upon Continuous-Time Markov Chains (CTMCs) (McQuarrie, 1967; Bartholomay, 1958), have been successfully introduced to overcome the modelling limitations of continuous methods. Stochastic systems are formally represented through a chemical master equation and have also been studied, rather directly, in the form of autocatalytic reactions systems, with approaches such as the one described in Dauxois et al. (2009).

However, since the analytic solution of the underlying equation is often infeasible for real size systems, these models are usually studied resorting to simulation approaches, mostly based on (variants of) Gillespie's stochastic simulation algorithm (Gillespie, 1977). Unfortunately, in some cases, even numerical simulation can be computationally very expensive. A compromise between accuracy and efficiency can be obtained by combining discrete and continuous evolution in so-called hybrid approaches (Pahle, 2009; Bortolussi and Policriti, 2009, 2013). In this context, we recall also (Salis and Kaznessis, 2005; Salis et al., 2006), where hybrid approaches for stochastic simulation of gene networks have been developed.

In this work, we consider a simple model of gene regulation in a transcription/translation genetic network, where a transcriptional repressor negatively regulates its own expression. This model has been widely studied (e.g. Marquez Lago and Stelling, 2010; Stekel and Jenkins, 2008), because it is a minimal system that explicitly describes the processes of transcription and translation and because it is a basic component of many complex biological systems. Despite its apparent simplicity, understanding its behaviour is not easy, because this is governed in a non-trivial way by several quantitative parameters. Actually, different parameter combinations lead to a range of qualitatively different dynamics. In particular, in Marquez Lago and Stelling (2010), the intensity of noise in this system is analysed in terms of various parameters, with special emphasis on the strength of the negative feedback. That paper demonstrates how the application of engineering principles to the role of feedbacks in a biological context could be misleading. The authors show, indeed, that noise generally increases with feedback strength, in contrast to the common knowledge. They also relate the possible different dynamical regimes with the different regions of the parameter space. The overall behaviour emerges from the complex interaction between feedback strength and other parameters of the system, governing the dynamics of the protein and of the mRNA.

Starting from the analysis proposed in Marquez Lago and Stelling (2010), we investigate the model with two goals in mind: (i) identifying the *parameters that play a key role in the regulation* process, by systematically studying the impact of parameters variations on the global dynamics of the system. In this way, we establish a link between the parameter space and the observed temporal patterns, i.e. the diverse behavioural phenotypes; (ii) quantifying the *impact of each reaction* of the modelled system on the overall dynamics and on noise patterns. This allows us to identify those reactions that, having a minor influence on the global noise pattern, can be safely approximated in a deterministic fashion.

At a higher level and on a longer term, we aim at setting up a systematic strategy for correctly building hybrid models of biochemical systems. In such models, only the most relevant sources of noise will be represented via a fully detailed stochastic description.

The handy dimension of our model of gene regulation allows us to play with different models and techniques. On the one hand, we study a stochastic model of the negative feedback loop to construct an exact picture of its possible behavioural patterns and of the effects of noise. This precision comes with a high computational cost, especially for certain parameter sets. On the other hand, we systematically apply various forms of model abstraction, in order to mitigate the inefficiency of exact stochastic simulation methods. In particular, we abstract the discrete stochastic dynamics into