Technical Report CoSBi 05/2008

Query-based Verification of Biochemical Oscillations through Probabilistic Model Checking

Paolo Ballarini

The Microsoft Research - University of Trento
Centre for Computational and Systems Biology

ballarini@cosbi.eu

Ivan Mura

The Microsoft Research - University of Trento
Centre for Computational and Systems Biology

mura@cosbi.eu

Radu Mardare

The Microsoft Research - University of Trento
Centre for Computational and Systems Biology

mardare@cosbi.eu

This is the preliminary version of a paper that will appear in the proceedings of the International Workshop From Biology To Concurrency and back (FBTC 2008), satellite workshop of ICALP 2008, July 12, 2008, Reykjavik, Iceland.
ENTCS 194(3).
available at DOI:10.1016/j.entcs.2009.02.002

Abstract

There are two popular approaches to modelling and verification of biological systems. In the first one a system of *ordinary differential equations* (ODEs) is developed and solved through numerical integration techniques yielding a *deterministic* solution of the model. In the second one a stochastic (discrete-state) model is obtained and analysed through (stochastic) simulation techniques, yielding statistical estimates of the system behaviour. We propose an alternative approach based on probabilistic/stochastic model checking techniques. Relaying on temporal logic formalisms we investigate the possibility for a query language for stating biological properties of a system's model. Such queries are then automatically evaluated through existing model checking algorithms. To exemplify such approach we consider a biochemical oscillator taken from the literature. First we study such oscillator through the classical ODEs and stochastic simulation approaches. Then we develop a number of (probabilistic) queries that we verify against a Markov chain model of the oscillator, through the probabilistic model checler PRISM.

1 Introduction

The quantitative analysis of the dynamics of biological systems is a fundamental task in *systems biology*, a new research field that focuses on the systematic study of complex interactions using an integrative approach rather than a reductive one [8]. Even simple biochemical networks of interacting proteins can show surprisingly complex behavior, a behavior that cannot be understood looking at the evolution of the single components but that instead requires a systemic analysis approach.

Traditionally, the study of time-dependent dynamics of biological systems has been addressed with deterministic approaches, based on ordinary differential equation (ODE, hereafter) models. ODEs provide quite an adequate abstraction for capturing the interactions and transformation of biochemical species, and come with a consolidated set of computational tools for model definition and solution [6]. More recently, stemming from the foundational work of Gillespie [4], a discrete approach to the modeling of biochemical systems has been gaining consensus, especially to model systems in which the continuous approximation does not seem to be justified, for instance when only a few molecules exist for some species.

The relationships between the continuous deterministic and the discrete stochastic representation of a given system are quite interesting and still subject of research. When the number of molecules that constitute a molecular network is low, the stochastic modeling may represent a more suitable tool to represent and analyze the dynamics of the system. On the other hand, as the number of molecules grows, abstracting discrete number of molecules into continuous concentration levels and representing evolution of dynamics

through a system of coupled ODEs provides very accurate representations and also has the advantage of not suffering from the state-space explosion problem that plagues stochastic modeling tools. Moreover, stochastic models are mostly solved via simulation, which may require performing a substantial number of simulation runs to compute statistically relevant results.

When dealing with biological systems that exhibit complex patterns of behavior, it may be difficult to check whether the dynamics of the system satisfies some interesting properties. One of such properties, quite often vaguely defined, is the oscillatory behavior that characterizes a variety of biological systems [2].Whereas it is comparatively easy to hypothesize the presence of oscillations through visual inspection of simulated time courses, another matter is to precisely verify whether the oscillations will continue forever or rather will definitely stop, either by being progressively damped down or abruptly interrupted. In this respect, the abstraction adopted by ODE models has some advantages, because the deterministic approach ensures that any periodic evolution will indefinitely repeat. However, as we will see in this paper, such properties may appear to hold only because of the continuous approximation. On the other hand, results obtained from discrete stochastic models can still provide clues about the existence of oscillatory behaviors, but the noise introduced by stochastic fluctuations may make them appear less pronounced. Furthermore, when oscillations do exist but their phase, period and amplitude is also varying stochastically, a multi-run analysis of simulation traces over the time domain does not help in characterizing the long-range behavior of the system.

In this paper we consider two examples of models of biochemical systems taken from the literature, which we use as case studies to demonstrate the advantages of a formal approach to the definition and verification of properties of biological systems, including oscillatory behaviors. The major contribution of this paper is the definition of an approach to the formal specification of properties, based on Probabilistic Computation Tree Logic (PCTL, hereafter) [5] and Continuous Stochastic Logic (CSL) [1]. PCT/CSL formulas can be used to precisely define properties of system dynamics, such as convergence of a variable to a fixed value, bound oscillation around a fixed value, perpetual alternation of growth and decrease phases, relationships among multiple variables. Such formulas define a query-language in which a user can encode a property of the systems he intends to verify. Probabilistic model checkers that accept PCTL formulas defined over the variables of a discrete stochastic model can then be used to verify whether the property holds of the system. In this paper, we show how the PRIMS [9] model checker can be used for this purpose. We apply our formal approach to the example models to show how system properties can be characterized and to demonstrate the insights that can be obtained into system dynamics with the verification with PRISM. We show the obvious advantages provided

by the exact quantitative solutions returned by a model-checking tool. In contrast, simulation of stochastic models can only provide statistically approximate solutions. Our proposed approach applies to general models of biological systems, its current limitation being the fact that only models having a bounded state space are amenable to analysis.

2 Conclusion

In this paper we have proposed a comparative study of modelling techniques for the analysis of biochemical systems. We have focused on a specific class of systems, ones whose behaviour is characterised by existence of oscillations in the quantities of reactants they consist of. First we have demonstrated how useful insight about an oscillating system can be gained by application of standard system biology's modelling techniques, such as solution of continuos deterministic (ODEs) model and simulation of a discrete stochastic model. However we have seen that both such methods lack to fulfil the modeller needs for analysis of the considered system. We have then argued that formal languages, such as temporal logic, may be applied to encoding relevant characteristic of an oscillator, providing a formalisation of the *oscillation permanence* in terms of two popular logic: CTL and PCTL. We have then developed a Markovian model of both a *transient* oscillator, known as the 3-way oscillator as well as of its *permanent oscillation* variant. We have demonstrated the effectiveness of automated temporal logic verification, by coding such models in the probabilistic model checker PRISM. we have been able to perform a thorough probabilistic analysis of several reachability as well as reward based logical formulae, which provided us with interesting outcomes which could not be obtained otherwise. Future developments of this work include the extension of the proposed methodology to the verification of more complex biological systems, such as the cell-cycle, a model of which is currently being developed and analysed.

2.1 Related work

Application of model checking techniques to system biology has been proposed in some other works. In [7], Kwiatkowska *et al.* developed a non trivial stochastic model of the complex Fibroblast Growth Factor (FGF) signalling pathway. They used PRISM to code the FGF pathway and developed a number of probabilistic properties, some of which reward-based, to verify relevant properties of the signalling mechanism. In [3] Gilber *et al.* propose a general overview about formal modelling techniques suited to the verification of biological systems. Based on Petri Net model of the ERK signal transduction pathway, they identify what type of formalisms/verification-technique, is suitable to what type of analysis the modeller is interested in.

Hence in the discrete-state modelling framework non-probabilistic model checking (i.e. standard temporal logic) has to be used to perform *qualitative analysis*, whereas probabilistic/stochastic is needed for *quantitative analysis*. Finally in the continuous-state modelling framework, ODEs solution as well as Linear Time Logic with constraints (LTLc) are to be used for model analysis.

References

[1] C. Baier, B. Haverkort, H. Hermann, and J.-P. Katoen. Model-checking algorithms for continuous-time Markov chains. *IEEE Trans. on Software Eng.*, Vol. 29(6):pp. 524–541, June 2003.

[2] M. Cavaliere and S. Sedwards. Modelling cellular processes using membrane systems with peripheral and integral proteins. In Springer, editor, *Proceedings of the fourth International Conference on Computational Methods in Systems Biology (CMSB2006), LNBI-LNCS 4210*, 2006.

[3] D. Gilbert, M. Heiner, and S. Lehrack. A unifying framework for modelling and analysing biochemical pathways using petri nets. In *In proceedings CMSB 2007, (Computational Methods in Systems Biology), Editors: M.Calder and S.Gilmore, Published by Springer-Verlag in LNCS/LNBI*, 2007.

[4] G. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem.*, 81(25):2340–2361, 1977.

[5] H. A. Hansson and B. Jonsson. A framework for reasoning about time and reliability. In *Proc. 10th IEEE Real -Time Systems Symposium*, pages 102–111, Santa Monica, Ca., 1989. IEEE Computer Society Press.

[6] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI — a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–3074, 2006.

[7] John Heath, Marta Kwiatkowska, Gethin Norman, Dave Parker, and Oksana Tymchyshyn. Probabilistic model checking of complex biological pathways. *Theoretical Computer Science*, 391:239–257, 2008.

[8] H. Kitano, editor. *Foundations of Systems Biology.* MIT Press, 2001.

[9] M. Kwiatkowska, G. Norman, and D. Parker. Probabilistic symbolic model checking with probabilistic symbolic model checking with PRISM. In J. Katoen and P. Stevens, editors, *Proceedings of the 8th International Conference on Tools and Algorithms for Construction and Analysis of Systems (TACAS 2002)*, volume 2280 of *Lecture Notes in Computer Science*, 2002.