

MIT Open Access Articles

A critical review on modelling formalisms and simulation tools in computational biosystems

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Machado, Daniel et al. "A Critical Review on Modelling Formalisms and Simulation Tools in Computational Biosystems." Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living. Ed. Sigeru Omatu et al. LNCS Vol. 5518. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. 1063–1070.

As Published: http://dx.doi.org/10.1007/978-3-642-02481-8_161

Publisher: Springer Berlin / Heidelberg

Persistent URL: <http://hdl.handle.net/1721.1/73902>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike 3.0



A critical review on modelling formalisms and simulation tools in Computational Biosystems

Daniel Machado¹, Rafael S Costa¹, Miguel Rocha², Isabel Rocha¹, Bruce Tidor³, and Eugénio C Ferreira¹

¹ IBB-Institute for Biotechnology and Bioengineering/Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
{dmachado, rafacosta, irocha, ecferreira}@deb.uminho.pt

² Department of Informatics/CCTC, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
mrocha@di.uminho.pt

³ Department of Biological Engineering/Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
tidor@mit.edu

Abstract. Integration of different kinds of biological processes is an ultimate goal for whole-cell modelling. We briefly review modelling formalisms that have been used in Systems Biology and identify the criteria that must be addressed by an integrating framework capable of modelling, analysing and simulating different biological networks. Aware that no formalism can fit all purposes we realize Petri Nets as a suitable model for Metabolic Engineering and take a deeper perspective on the role of this formalism as an integrating framework for regulatory and metabolic networks.

1 Introduction

Systems Biology is a new field of science with the goal of analysing the most complex biological processes by looking at the cell as a whole [17]. The traditional study of individual biological components is far too limited. Looking at metabolic networks, there are thousands of reactions with rates dependent not only on kinetic parameters but also on enzyme concentrations. The later are consequence of the functioning of regulatory networks, and to account for the dependency among biological processes, we must also take into account the role of signalling pathways in the control of the most fundamental cellular processes.

Integration of different types of biological networks [8] is fundamental to the purpose of Systems Biology and can bring major benefits in many research efforts, such as drug development and biotechnological production processes. This would benefit from the creation of a common modelling framework that takes into account different *entities* (genes, proteins, metabolites, *etc*) and *relationships* (metabolic reactions, interactions, regulation, transport, *etc*).

Metabolic Engineering [32] is concerned with *how* genotype translates into phenotype, and *what* perturbations can be made into the genotype to obtain a

desired phenotype for industrial applications. Several problems and requirements arise for this kind of study, such as how to deal with incomplete information, how to manipulate large models, how to extract valuable information from network topology, how to analyse these models, and how to predict behavioural changes by gene addition/removal or up/down-regulation. Pure mathematical models alone can hardly fulfil such demands, which reflects the need to search for suitable computational models.

This work addresses the study of computational formalisms for the representation of integrated biological networks and the identification of the features that a suitable framework should present. Learning from the experience in software engineering it is obvious that formal methods are vital to the implementation of such complex computational systems. Moreover, although the cell can be seen as a computational system, unlike most common ones, it is massively concurrent with thousands of interactions occurring simultaneously. Therefore, formalisms coming from the foundations of concurrency theory may be specially attractive.

The next section reviews modelling formalisms currently used in Systems Biology. Section 3 enumerates the criteria used to select an appropriate formalism and Section 4 explores the advantages of using Petri Nets as models for these systems. Section 5 presents our conclusions and future directions.

2 Modelling Formalisms in Systems Biology

Due to its multidisciplinary nature, many modelling formalisms have been used in Systems Biology (for reviews see [11, 18]).

Mathematical models such as nonlinear ordinary differential equations (ODEs) have widespread use in all kinds of biological processes, as for example, in metabolic pathways [5]. ODEs can accurately model complex systems but they are dependant on a large number of parameters difficult to measure experimentally. Moreover, they usually do not have closed-form solutions and need to be solved numerically. S-systems [27] are a canonical power-law formulation that work as an approximation to ODEs, but are faster to compute and have qualitative solutions. Nevertheless, they still require the estimation of a large number of parameters. Other types of differential equations, such as stochastic differential equations (SDEs) and partial differential equations (PDEs) can be used to include, respectively, stochastic effects and spatial distribution. These type of mathematical models are the basis for whole-cell simulation software such as E-Cell [33].

Computational models are usually more flexible than purely mathematical ones and have been used to model biological systems at different levels of abstraction.

Boolean networks are used to represent gene regulatory networks [16]. In these models, genes are represented by nodes with active or inactive states, connected by logical operators that represent activation and inhibition patterns. These networks can be inferred from gene expression levels (*e.g.* microarray data). Depending on the available data about the system and the level of ab-

straction, other models can be used for the same purpose such as Bayesian networks, Fuzzy logic and Markov chains [14].

Petri nets are formal graphical representations used to model complex concurrent systems. They are bipartite graphs containing two kinds of nodes, places and transitions, connected by directed arcs. Although originally based on discrete events, different extensions have evolved such as continuous, stochastic, hybrid and high-level nets. Due to this flexibility they have widespread use in Systems Biology for modelling all types of biological networks [22, 4, 13].

Process algebras are a family of formal languages for modelling concurrent systems. They describe systems in terms of compositions of communicating processes. π -calculus is used in [26] to model the RTK-MAPK pathway. Extensions to π -calculus include κ -calculus used to model protein interactions [9], and SpacePi to model molecular processes accounting for spatial distribution [15]. The Ambient calculus process algebra is used in [25] to model molecular processes with support for cellular compartments.

Rule-based systems such as BioNetGen [1] can naturally model biochemical reactions involving species with many activation states avoiding the combinatorial complexity that arises in these cases. BIOCHAM [3] is a tool that provides a rule-based language to write models of biochemical networks, and a temporal logic language for analysing their properties.

Statecharts are a formal visual language for description of reactive systems. They are used in [7] to model and simulate cellular differentiation using a technique called *reactive animation*. The modular and hierarchical structure of statecharts makes them specially suitable for multi-scale modelling. Using a bottom-up approach it is possible to model biological systems from molecular interactions to cells, organisms and even populations. These models can explain how emergence at a scale arises from interactions at lower scales.

Hybrid automata are dynamic systems with continuous variables ruled by differential equations and discrete state transitions. They are used in [12] to model the Delta-Notch signalling pathway. The authors present an algorithm for computing the initial states *backward reachable* from a steady-state. The key advantage being that both modelling and analysis are symbolic, *i.e.* no parameters need to be numerically instantiated.

3 Criteria for an adequate formalism

Although complementary in their functionality, none of the modelling formalisms presented in the previous section is suitable to model the complexity of biological processes as a whole. An unifying formalism able to express all the structure and functionality of cellular processes should follow several criteria that will be roughly enumerated.

- **Integrative:** A common modelling framework should integrate different kinds of biological processes. ODEs are used to model all kind of phenomena. Petri nets have been used to model metabolic, signalling and regulatory networks [22]. π -calculus has also been used in different kinds of processes, as

in protein interactions [9], signalling pathways [26] and molecular processes [15]. Models should be able to be represented in a standard format. Some features proposed for future versions of SBML [10] are coincident with extensions already developed for some formalisms (*e.g.* composition of models from submodels, spatial features, diagram layouts).

- **Intuitive:** Systems Biology is a multidisciplinary research field that gathers biologists, computer scientists and engineers. Therefore, models should ideally be expressed as intuitively as possible and easily interpreted by people from different areas. For that matter, graph and diagram-based formalisms such as petri nets and statecharts can be visually appealing and similar to common informal notations.
- **Scalable:** The life-cycle of a model is an iterative process of experimentation and successive model refinement with experimental data. From the cell as a black-box interacting with its environment, to a fully detailed biochemical network, different levels of abstraction should be supported. Modelling approaches in systems biology are usually classified as *top-down*, *bottom-up* or *middle-out* [21]. Moreover, support for different time scales and compartments are important features. BioAmbients [25] is an example of a process algebra that expresses compartments.
- **Qualitative:** Part of the studies in Systems Biology are related to the structural and topological properties that emerge from the complex organization of biological networks. Dynamic models based on ODEs can accurately simulate the dynamic behaviour of such networks but fail to express their structure and modularity. Graph-based representations however, can successfully capture such properties.
- **Quantitative:** A model should be able to simulate cellular behaviour and response to perturbations. Metabolic steady-state models can be simulated using stoichiometric information and an optimization criterion as in flux balance analysis. Dynamic models are usually based on ODEs and simulations are performed by numerical integration. Novel modelling formalisms should consider the work already developed for analysis of mathematical models. This has been explored in [2] where a method for automatically deriving ODEs from process algebra models is presented.

4 Petri Nets

Petri nets have been widely used as a modelling language for different applications such as concurrent programming, workflow management, manufacturing process design and communication systems. Based on the criteria defined, petri nets sound the most promising formalism for the goals of our work. They are a graphical and mathematical formalism, therefore intuitive and amenable to analysis. Their application for modelling biological processes began in 1983 [24], where the authors addressed the limitation of quantitative modelling due to lack of experimental data and proposed them as a framework for qualitative modelling of metabolic pathways. They illustrate an example of the fructose pathway and

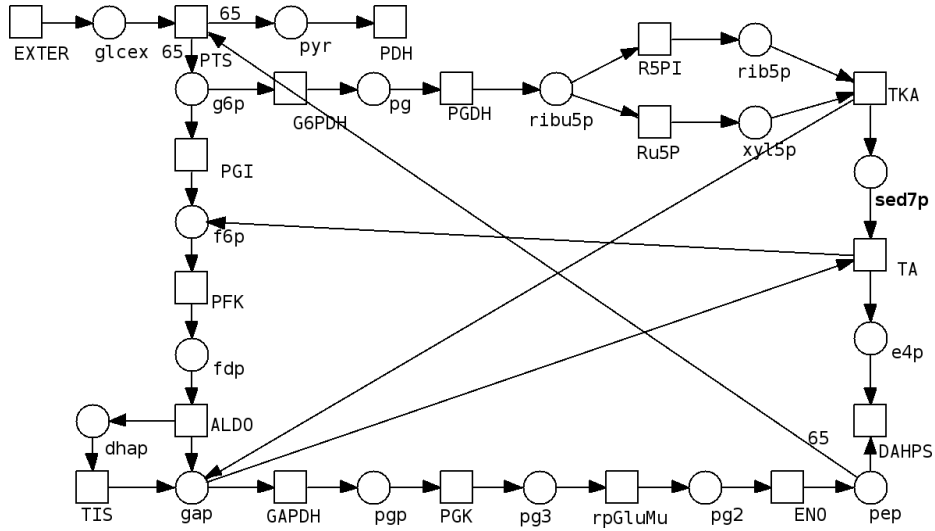


Fig. 1. Petri net model of one of the elementary flux modes of the central carbon metabolism model of *E. coli*.

introduce the biological meaning of some properties such as *liveness*, *reachability* and *reversibility*. Also, they present concepts like net reduction and abstraction, whereby a transition may be an abstraction of a more complex subnet (*e.g.* enzyme catalyzed reaction).

There is currently a considerable amount of literature on the application of petri nets to Systems Biology [22, 4, 13]. The biological meaning of several properties has been explored as the basis of model analysis [24, 34, 13]. For example *p-invariants*, *t-invariants*, *siphons* and *traps* represent, respectively, conservation, elementary flux modes, un replenished resources and compound accumulation.

Extensions to petri nets have evolved in different directions, including features like inhibitory arcs, timing, stochastic effects and continuous transitions. In [13] the authors explore how discrete qualitative models can be extended to quantitative models using a stochastic or continuous approach and how the three paradigms complement each other. A petri net model of the ERK signal transduction pathway is used as a case-study. Model checking techniques are applied to all three approaches using different types of temporal logic.

Petri nets are suitable for metabolic engineering purposes as they provide an integrating framework for regulatory and metabolic networks. A methodology for converting boolean regulatory networks to petri nets is presented in [31]. An integrated regulatory and metabolic petri net model of tryptophan biosynthesis in *E. coli* is developed and analysed in [29]. In [34] the authors discuss the topological analysis of qualitative metabolic models based on petri nets and highlight the advantages of this framework for studying knockout mutations.

Quantitative models must take into account the discrete nature of regulatory networks and the continuous nature of metabolic pathways. Hybrid models are

therefore required for quantitative simulation of regulated metabolic networks. In [6] hybrid petri nets are used to model the urea cycle regulated metabolic network. Hybrid Functional Petri Nets (HFPNs) are introduced in [19] and used to model and simulate the *E. coli* glycolytic pathway with *lac* operon regulation.

Interchange between petri net models and the SBML standard is essential for support of available models and tools. This has been addressed in [28] where the authors developed a prototype tool that translates SBML to the Petri Net Markup Language (PNML).

Thus far, no single tool supports all petri net extensions. The CPN Tools software supports coloured petri nets and has industrial use in many control applications [23]. Genomic Object Net was developed to support the HFPN specification [20], this tool is now known as Cell Illustrator and requires a commercial license. Snoopy [13] is a general purpose tool for design and simulation of graph-based formalisms and is free for non-commercial use. It supports different petri net extensions (extended, timed, stochastic, continuous), and features an SBML import wizard.

We tested Snoopy by importing an SBML model of the dynamic central carbon metabolism of *E. coli* described in [5]. When importing into a discrete petri net only stoichiometric information is considered and reversible reactions must be separated into two transitions. This model can be exported for analysis by other tools. Exporting to the INA tool [30] allowed us analyse several net properties. We then extended our model by importing it into a continuous petri net. Snoopy successfully performed dynamic simulations of the metabolism by numerical integration of the continuous model. Finally, we compared the semi-positive *t-invariants* (elementary flux modes) computed by INA with the flux distribution of the dynamic model at steady-state and realized that this model operates closely to the elementary mode shown in Fig. 1. This is a simple test-case, yet it illustrates the interplay between modelling, analysis and simulation.

5 Conclusions and Future Work

The diversity of problems to be solved in Systems Biology gave rise to approaches using many different paradigms. We briefly reviewed modelling formalisms that have been used to model biological processes and classified the fundamental features required by a formalism suitable to integrate all major kinds of biological networks. Petri nets are an intuitive graphical formalism with a sound mathematical foundation. Extensions such as stochastic, continuous and hybrid nets provide a great flexibility for modelling metabolic, regulatory and signalling networks. Making use of hierarchical features for zooming between different levels of abstraction may become specially useful for handling and comprehending large scale models. Tool support for design and analysis of petri net models, particularly for biological context is still rather disperse. This is an area where standardization and an open source collaboration initiative could have great added value. Our future developments will use petri nets as the basis of a framework

for modelling and analysis of integrated regulatory and metabolic networks. This framework will improve the study of *E. coli* for industrial applications.

Acknowledgments. Research supported by Phd grant SFRH/BD/35215/2007 from the Fundação para a Ciência e a Tecnologia (FCT) and the MIT-Portugal program.

References

1. M.L. Blinov, J.R. Faeder, B. Goldstein, and W.S. Hlavacek. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics*, 20(17):3289–3291, 2004.
2. M. Calder, S. Gilmore, and J. Hillston. Automatically deriving ODEs from process algebra models of signalling pathways. *Computational Methods in Systems Biology 2005*, pages 204–215, 2005.
3. L. Calzone, F. Fages, and S. Soliman. BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22(14):1805–1807, 2006.
4. C. Chaouiya. Petri net modelling of biological networks. *Briefings in Bioinformatics*, 8(4):210, 2007.
5. C. Chassagnole, N. Noisommit-Rizzi, J.W. Schmid, K. Mauch, and M. Reuss. Dynamic modeling of the central carbon metabolism of Escherichia coli. *Biotechnology and Bioengineering*, 79(1):53–73, 2002.
6. M. Chen. Quantitative Petri Net Model of Gene Regulated Metabolic Networks in the Cell. *In Silico Biology*, 3(3):347–365, 2003.
7. I.R. Cohen and D. Harel. Explaining a complex living system: dynamics, multi-scaling and emergence. *Journal of The Royal Society Interface*, 4(13):175–182, 2007.
8. M.W. Covert, N. Xiao, T.J. Chen, and J.R. Karr. Integrating Metabolic, Transcriptional Regulatory and Signal Transduction Models in Escherichia coli. *Bioinformatics*, 2008.
9. V. Danos and C. Laneve. Formal molecular biology. *Theoretical Computer Science*, 325(1):69–110, 2004.
10. A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochem. Soc. Trans*, 31(Pt 6):1472–3, 2003.
11. J. Fisher and T.A. Henzinger. Executable cell biology. *Nature Biotechnology*, 25(11):1239, 2007.
12. R. Ghosh and C. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-notch protein signalling. *Systems Biology, IEE*, 1(1):170–183, 2004.
13. M. Heiner, D. Gilbert, and R. Donaldson. Petri Nets for Systems and Synthetic Biology. *Formal Methods for Computational Systems Biology: 8th International School on Formal Methods for the Design of Computer, Communication, and Software Systems, Sfm 2008 Bertinoro, Italy, June 2-7, 2008*, 2008.
14. T. Ideker and D. Lauffenburger. Building with a scaffold: emerging strategies for high-to low-level cellular modeling. *Trends in Biotechnology*, 21(6):255–262, 2003.
15. M. John, R. Ewald, and A.M. Uhrmacher. A Spatial Extension to the π Calculus. *Electronic Notes in Theoretical Computer Science*, 194(3):133–148, 2008.

16. SA Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol*, 22(3):437–67, 1969.
17. H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, 2002.
18. W. Materi and D.S. Wishart. Computational systems biology in drug discovery and development: methods and applications. *Drug Discovery Today*, 12(7-8):295–303, 2007.
19. H. Matsuno, Y. Tanaka, H. Aoshima, A. Doi, M. Matsui, and S. Miyano. Biopathways Representation and Simulation on Hybrid Functional Petri Net. *In Silico Biology*, 3(3):389–404, 2003.
20. M. Nagasaki, A. Doi, H. Matsuno, and S. Miyano. Genomic Object Net: I. A platform for modelling and simulating biopathways. *Applied Bioinformatics*, 2(3):181–184, 2003.
21. D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3(6):459–463, 2002.
22. JW Pinney, DR Westhead, and GA McConkey. Petri Net representations in systems biology. *Biochem. Soc. Trans*, 31:1513–1515, 2003.
23. A.V. Ratzert, L. Wells, H.M. Lassen, M. Laursen, J.F. Qvortrup, M.S. Stissing, M. Westergaard, S. Christensen, and K. Jensen. CPN Tools for Editing, Simulating, and Analysing Coloured Petri Nets. *Lecture Notes in Computer Science*, pages 450–462, 2003.
24. V.N. Reddy, M.L. Mavrovouniotis, and M.N. Liebman. Petri Net Representations in Metabolic Pathways. In *Proceedings of the 1st International Conference on Intelligent Systems for Molecular Biology*, pages 328–336. AAAI Press, 1993.
25. A. Regev, E.M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. BioAmbients: an abstraction for biological compartments. *Theoretical Computer Science*, 325(1):141–167, 2004.
26. A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. In *Pac Symp Biocomput*, volume 459, page 70, 2001.
27. M.A. Savageau and E.O. Voit. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Mathematical Biosciences*, 87(1):83–115, 1987.
28. O. Shaw, A. Koelmans, J. Steggles, and A. Wipat. Applying Petri Nets to Systems Biology using XML Technologies. *ATPN 2004*, 2004.
29. E. Simao, E. Remy, D. Thieffry, and C. Chaouiya. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in E. Coli. *Bioinformatics*, 21(90002), 2005.
30. P.H. Starke. INA: Integrated Net Analyzer. *Reference Manual*, 1992.
31. LJ Steggles, R. Banks, and A. Wipat. Modelling and Analysing Genetic Networks: From Boolean Networks to Petri Nets. *Computational Methods in Systems Biology*, 2003.
32. G. Stephanopoulos. Metabolic engineering. *Biotechnology and Bioengineering*, 58, 1998.
33. M. Tomita, K. Hashimoto, K. Takahashi, T.S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J.C. Venter, et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics*, 15(1):72–84, 1999.
34. I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, 3:0029, 2003.