Proceedings of
The Évry Spring School
on

Modelling Complex Biological Systems in the Context of Genomics

May 3rd - 7th, 2010

Edited by

Patrick Amar, François Képès, Vic Norris

"But technology will ultimately and usefully be better served by following the spirit of Eddington, by attempting to provide enough time and intellectual space for those who want to invest themselves in exploration of levels beyond the genome independently of any quick promises for still quicker solutions to extremely complex problems."

Strohman RC (1977) Nature Biotech 15:199

FOREWORD

What are the salient features of the new scientific context within which biological modelling and simulation will evolve from now on? The global project of high-throughput biology may be summarized as follows. After genome sequencing comes the annotation by 'classical' bioinformatics means. It then becomes important to interpret the annotations, to understand the interactions between biological functions, to predict the outcome of perturbations, while incorporating the results from post genomics studies (of course, sequencing and annotation do not stop when simulation comes into the picture). At that stage, a tight interplay between model, simulation and bench experimentation is crucial. Taking on this challenge therefore requires specialists from across the sciences to learn each other's language so as to collaborate effectively on defined projects.

Just such a multi-disciplinary group of scientists has been meeting regularly at Genopole, a leading centre for genomics in France. This, the *Epigenomics project*, is divided into five subgroups. The *GolgiTop* subgroup focuses on membrane deformations involved in the functionning of the Golgi. The *Hyperstructures* subgroup focuses on cell division, on the dynamics of the cytoskeleton, and on the dynamics of *hyperstructures* (which are extended multi-molecule assemblies that serve a particular function). The *Observability* subgroup addresses the question of which models are coherent and how can they best be tested by applying a formal system, originally used for testing computer programs, to an epigenetic model for mucus production by *Pseudomonas aeruginosa*, the bacterium involved in cystic fibrosis. The *Bioputing* group works on new approaches proposed to understand biological computing using computing machine made of biomolecules or bacterial colonies. The *SMABio* subgroup focuses on how multi-agents systems (MAS) can be used to model biological systems.

The works of subgroups underpinned the conferences organised in Autrans in 2002, in Dieppe in 2003, in Evry in 2004, in Montpelliers in 2005, in Bordeaux in 2006, back to Evry in 2007, in Lille in 2008 and in Nice in 2009. The conferences in Evry in 2010 which as reported here, brought together over a hundred participants, biologists, physical chemists, physicists, statisticians, mathematicians and computer scientists and gave leading specialists the opportunity to address an audience of doctoral and post-doctoral students as well as colleagues from other disciplines.

This book gathers overviews of the talks, original articles contributed by speakers and subgroups, tutorial material, and poster abstracts. We thank the sponsors of this conference for making it possible for all the participants to share their enthusiasm and ideas in such a constructive way.

Patrick Amar, Gilles Bernot, Marie Beurton-Aimar, Bruno Goffinet, Eric Goles, Janine Guespin, Jürgen Jost, Marcelline Kaufman, François Képès, Pascale Le Gall, Reinhard Lipowsky, Jean-Pierre Mazat, Victor Norris, El Houssine Snoussi.

ACKNOWLEDGEMENTS

We would like to thank the conference participants, who have contributed in a way or another this book. It gathers overviews of the talks, discussions and roundtables, original articles and tutorial material contributed by speakers, abstracts from attendees, posters and lectures proposed by the epigenesis groups to review or illustrate matters related to the scientific topic of the conference.

Of course the organisation team would like to express gratitude to all the staff of the Évry *All Seasons Hotel* for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows the *Faculté des Métiers* Copyright *Ville d'Évry*.

We would also like to express our thanks to the sponsors of this conference for their financial support allowing the participants to share their enthusiasm and ideas in such a constructive way.
They were:

- Centre National de la Recherche Scientifique (CNRS):
  http://www.cnrs.fr
- Genopole® Evry:
  http://www.genopole.fr
- GDRE CNRS 513 Biologie Systémique:
  http://www.mpi-magdeburg.mpg.de/CNRS_MPG
- Consortium BioIntelligence (OSEO)
- PRES UniverSud Paris:
  http://www.universud-paris.fr/
- Institut National de Recherche en Informatique et en Automatique (INRIA):
  http://www.inria.fr/
- GDR CNRS 3003 Bioinformatique Moléculaire:
  http://www.gdr-bim.u-psud.fr
- Réseau national des systèmes complexes (RNSC):
  http://rnsc.csregistry.org/tiki-index.php
- INRA MIA:
  http://www.inra.fr
- Fondation Scientifique Fourmentin-Guilbert:
  http://www.fourmentinguilbert.org

THE EDITORS

INVITED SPEAKERS

TATSUYA AKUTSU	Kyoto Univ., (Japan)
JULIO ARACENA	Univ. Concepcion, (Chili)
ESHEL BEN JACOB	Tel-Aviv Univ., (IL)
JEAN-PAUL COMET	Univ. Nice, (F)
VINCENT FROMION	INRA, Jouy-en-Josas, (F)
VASSILY HATZIMANIKATIS	EPF Lausanne, (CH)
JACQUES VAN HELDEN	ULB, Brussels, (BE)
NANCY KLECKNER	Harvard Univ., MA, (USA)
MELANIE MUELLER	MPIKG Postdam, (DE)
FRANÇOIS NÉDÉLEC	EMBL, Heidelberg, (DE)
ERIK VAN NIMWEGEN	Biozentrum Basel, (CH)
GERMÁN RIVAS	Centro de Investigaciones Biológicas, Madrid (SP)
ERICH SACKMANN	Technische Universität, München, (DE)
JÖRG STÜLKE	Georg-August-Universität, Göttingen, (DE)
JEAN WEISSENBACH	Institut de génomique, CEA, Evry, (F)

CONTENTS

PART I INVITED TALKS

Modeling cytoskeletal structures with cytosim

François Nédélec[1]

[1] EMBL, Heidelberg, Germany

Abstract

Living cells have a system of fibers and associated proteins, called the cytoskeleton, which provides the mechanical support necessary for migration, polarization, division, etc. We study the cytoskeleton to better understand how an initially uniform set of proteins can collectively generate order. Indeed, the monomers can spontaneously assemble to form fibers, and multiple fibers can self-organize into higher order structures spanning the entire cell. The cytoskeleton offers many examples of remarkable structures generated by the uncoordinated interactions of its constituant proteins. The self-organization can sometimes be reconstituted in vitro, using a defined and limited set of components. In this talk, we will discuss a complementary approach, in which one uses numerical simulations to study the emergent properties. We will introduce cytosim, ang give some examples of how this simulation can be configured by a user. We will also give an example of current investigation in the field of mitosis research.

Cytomimetic reconstitution of the bacterial proto-ring complex in the test tube

German Rivas[1]

[1] CIB / CSIC, Madrid, Spain

Abstract

Our long-term experimental goal is to build the minimum multi-protein complex required to initiate bacterial cell division in the test tube. The dynamic division ring formed at mid-cell towards the end of the cell cycle is composed by at least ten division specific proteins that interact in a reversible manner to form the functional assembly. The first multi-protein complex formed is the proto-ring that initiates division. In E. coli the proto-ring is a complex of three proteins (FtsZ, FtsA, and ZipA) assembling on the cytoplasmic membrane, which is required for the incorporation of the remaining proteins at the mature ring. The GTP-mediated assembly and disassembly of FtsZ (the bacterial ancestor of the eukaryotic tubulin) are thought to be essential for the formation of the septal ring.

We are using a complementary biophysical, biochemical, genetic and imaging approach to reconstitute the proto-ring components into a variety of biomimetic membrane systems, including nanodiscs and vesicles. To provide a native cell environment, these studies are being done in a cytomimetic medium to reproduce the crowded intracellular environment, physiological osmolarity and energy supply pools. This set of tools will define the physicochemical conditions that modulate the energetics and dynamics of FtsZ association to the cytoplasmic membrane and its eventual dissociation from it.

Mechanical forces control the structure and function of cells by stimulating genetic expressions and biochemical signalling

Erich Sackmann[1]

[1] Physics Department Technical University Munich James Franck Str.1 D85747 Garching, DE

Abstract

The physics of cells is greatly stimulated by the growing evidence that mechanical forces play a ubiquitous role for the adaption of the material properties of cells to their biological function. Cells can sense external forces through their composite cell envelope and react in an interactive way either rapidly, through the stimulation of biochemical switches, or slowly through stimulation of genetic expressions. After an introduction into basic physical properties of the composite envelope and the cytoplasmic space of cells several examples of force controlled processes are discussed:

i Stem cells can differentiate by matching of their mechanical impedance to that of the environment.

ii Cells exhibit stress and strain sensors and can adapt the adhesion strength to external forces.

iii Cells can move over surfaces by propagation of adhesion-induced microdomains which stimulate solitary actin gelation waves through signalling lipids (PI-3,4,5-P3).

Discrete and Hybrid Modeling of Gene Regulatory Networks

Jean-Paul Comet[1]

[1] Laboratoire I3S, UMR 6070 CNRS/UNSA,Algorithmes-Euclide-B,
2000 route des Lucioles, B.P. 121, F-06903 Sophia-Antipolis, France

Abstract

Computational modeling of genetic regulatory networks aims at deep understanding of how their components are controlled, thus allowing the prediction of a set of non-obvious conclusions that can be experimentally tested. While data on the connectivity among elements of the network is becoming increasingly available, kinetic data of the associated biochemical reactions remain to be determined. This parameter identification problem constitutes the cornerstone of the modeling approaches. More precise the available information about the dynamics of the system, more precise can be the model. But if precision of the model is higher than the one of the knowledge of the biological system, the precision given by computer simulations is only a consequence of an arbitrary choice of parameter values.

This comment motivated some researchers to develop methods where this identification problem is tractable. In particular Rene' Thomas' discrete modeling of gene regulatory networks (GRN) is a well-known approach to study the dynamics resulting from a set of interacting genes. It deals with some discrete parameters which reflect the possible targets of trajectories. Those parameters are a priori unknown, but they may generally be deduced from a well-chosen set of biologically observed trajectories.

Besides, it neglects the time delay for a gene to pass from one level of expression to another one whereas information on the time mandatory for the system to go from a state to another one is often available. Such an information is then not useful in such a framework to face up to the parameter identification problem. It makes more useful the classes of models where time is explicit. We then present an hybrid extension of pure discrete approach of R. Thomas in which time is explicit: New parameters, i.e. delays mandatory for a gene to go from a discrete abstract level to another one, allow the determination of time along a trajectory. Such a modeling framework preserves powerful enough computer-aided reasoning capabilities. The identification problem seems to become more difficult because of the increased number of parameters but computer is able to reject a large class of parameters.

In this presentation, we sketch first the discrete framework introduced by R. Thomas, in which each gene has several pertinent abstract levels, and we illustrate how computer science technics are able to automate the determination of parameters compatible with available knowledge about the dynamics of the system. Then we introduce a possible extension of such a modeling in which time is explicit, and show how constraints on these parameters can be deduce from the differential framework.

Update digraph and dynamical behavior in Boolean networks

Julio Aracena[1]

[1] University of Concepcion, Chili

Abstract

Boolean networks (BNs) have been used as models of gene regulation and other biological networks. One key element in these models is the update schedule, which indicates the order in which states are to be updated. For a long time synchronous update was the default choice for BN researchers, in part because of the scarcity of actual models of real networks. In some cases, most of the update schedules yield a different dynamical behavior than the parallel one. We have studied the robustness of Boolean networks with respect to different deterministic update schedules (synchronous, block-sequential, sequential). For a given Boolean network, we define equivalence classes of update schedules with the same dynamical behavior, introducing a labeled graph, named update digraph, which helps to understand the dependence of the dynamics with respect to the update

In this talk, we will focus on the update digraph and the relation between its structural characteristics and the robustness of the dynamical behavior of a BN against changes in the update schedule. For instance, we will show how the update digraph associated to a BN is related to the number and size of the equivalence classes of update schedules yielding a same dynamical behavior. This enables us to roughly quantifier the number of different dynamics and the robustness of a network when we change the update schedule. Besides, we will exhibit necessary and sufficient conditions on the update digraph for the existence of different equivalence classes keeping a certain dynamical property (e.g. limit cycles) but not necessarily the whole dynamics.

Algorithmic Aspects of Analysis and Control of Boolean Networks

Tatsuya Akutsu[1]

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan

Abstract

Analysis of genetic networks is an important topic in bioinformatics and computational systems biology. For that purpose, various mathematical models of genetic networks have been proposed and utilized. In this talk, we focus on the Boolean network (BN) model. Furthermore, we focus on detection/enumeration of attractors and finding of control actions for BNs. We give a brief introduction of these problems and review algorithmic results on these problems with focusing on our works. For detection of attractors, we review an NP-hardness result, a recursive algorithm, and SAT-based algorithms. For control of BNs, we review NP-hardness results and a polynomial time algorithm for tree-structured networks. We also review practical integer linear programming-based algorithms for both problems and for the minimum knockout problem for Boolean models of metabolic networks. Finally, we discuss about possible future developments on these problems.

Motif Activity Response Analysis: Inferring genome-wide transcription regulation in mammals

Erik Van Nimwegen[1]

[1] Biozentrum Basel, CH

Abstract

I will discuss an integrated computational approach, called motif activity response analysis (MARA), for reconstructing transcription regulatory networks in mammals from genome-wide expression data. Based on deep sequencing data of transcription start sites we obtained a comprehensive 'promoteromes' in human and mouse, and using probabilistic comparative genomic methods we predict binding sites for over 200 regulatory motifs in proximal promoters genome-wide. Motif Activity Response Analysis (MARA) models genome-wide gene expression profiles in terms of these predicted regulatory sites and I will describe how MARA identifies, for a given system of study, the key regulators driving expression changes, their activity profiles across the samples, and the sets of target promoters of each regulator. Time permitting I will talk about how MARA can be extended to incorporate epigenetic changes to chromatin structure.

Organisation of metabolism in *Bacillus subtilis*: Evidence for the presence of protein complexes in central metabolism

Jörg Stülke[1]

[1] Georg-August-Universität, Göttingen, Germany

Abstract

Any cell contains a huge number of different proteins. For many organisms, including the Gram-positive soil bacterium *Bacillus subtilis*, these components are known due to the availability of genome information. However, only the interactions among the biological macromolecules and with the metabolites make this collection a living thing. We have studied such interactions for the central metabolic pathways of *B. subtilis*. Given the large amount of different pathways and reactions, enzymes and metabolites, the idea that enzymes of important pathways are not just dissolved in a cytoplasmic soup seems highly attractive. Such interactions were first identified by using the SPINE approach of *in vivo* cross-linking [1] and then confirmed by bacterial two-hybrid analyses. Indeed, we observed that glycolytic enzymes could contribute to many interactions in the cell. Very prominent among these interactions are those with a RNA processing machinery, the RNA degradosome, and those among different glycolytic enzymes themselves [2]. It seems that the glycolytic enzymes form a complex in *B. subtilis*. Similarly, enzymes of the tricarboxylic acid cycle interact with each other. The most prominent interactions are those between citrate synthase, isocitrate dehydrogenase, malate dehydrogenase, and fumarase. Our findings are in good agreement with the hypothesis that metabolism takes place in an organized and structured manner.

References

[1] Herzberg, Flórez Weidinger, Dörrbecker, Hübner, Stülke and Commichau (2007) *Proteomics* **7**: 4032-4035

[2] Commichau, Rothe, Herzberg, Wagner, Hellwig, Lehnik-Habrink, Hammer, Völker, and Stülke (2009) *Mol. Cell. Proteomics* **8**: 1350-1360

Recent progresses on the metabolism modeling of Bacteria: definition of local and global modules and a first explanation of their emergence

Vincent Fromion[1]

[1] INRA, Unité MIG, Jouy-en-Josas, France

Abstract

Growth is a highly optimized process in bacteria since the battle for the conquest of ecological niches is though. Understanding the key elements governing the growth rate management is then crucial to further understand and predict the bacterium behavior with respect to various environmental conditions. Constraint-based approaches integrating the whole metabolic network of an organism such as Flux Balance Analysis successfully predicted the maximum growth rate reachable in various conditions [1]. However, limitations exist since they fail to predict a number of commonly observed metabolic strategies.

Here we showed that the sharing of resources between the cellular processes intrinsically and structurally limits the growth rate [2]. We formalized the problem of resource repartition at the cell scale as a convex optimization problem and demonstrated the existence of a trade-off for the protein repartition between the translation apparatus and the metabolic network. Moreover, the resolution of this optimization problem for Bacillus subtilis allows the estimation for a given medium composition of:

i the maximal growth rate reachable;

ii the concentrations of ribosomes, the concentration of proteins involved in the metabolic network and more generally the resource repartition between cell components;

iii the flux distribution.

Besides, we also predicted the induction and repression of metabolic subsystems with respect to the environmental condition which correspond to the recently identified elementary modules of *Bacillus subtilis* [3]. Finally we also recovered the well-known evolution of ribosomes and metabolic proteins with respect to the growth rate of the Copenhagen school [4].

Another general conclusion of this work is the successful use of tools and methods based on convex optimization in biology. The formalization of the cell behavior is suitable for convex optimization and strong structural properties have been obtained allowing us to explain the emergence of functional

modules in the metabolic network regulation. The links between these two fields (biology and optimization) have to be strengthened in order to investigate fundamental questions such as the evolution of regulatory networks of organisms with respect to the ecological niche.

References

[1] Edwards J, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* **97(10)**: 5528-5533, 2000.

[2] Goelzer A, Fromion V, Scorletti G. Cell Design in Bacteria As a Convex Optimization Problem. *48th IEEE Conference on Decision and Control*, Shangai, China, 2009.

[3] Goelzer A, Bekkal Brikci F, Martin-Verstraete I, Noirot P, Bessières P, Aymerich S, and Fromion V, Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*, *BMC Systems Biology* **(2)** 1-20, 2008.

[4] Marr AG. Growth rate of *Escherichia coli*. *Microbiol Rev.* **55(2)**: 316-333, 1991.

To be announced

Jean Weissenbach[1]

[1] Institut de génomique, CEA, Evry, France

Abstract

To be announced

Vassily Hatzimanikatis[1]

[1] EPF Lausanne, Switzerland

Abstract

Finding relevant paths in the not-so-small world of metabolic networks

Jacques Van Helden[1]

[1] Service de Bioinformatique des Génomes et des Réseaux (BiGRe).
Université Libre de Bruxelles. Belgium

Abstract

Metabolic networks can be built by interconnecting all known reactions with their substrates and products. Since 10 years, metabolic and other biological networks have been claimed to present "universal" properties: power-law degree distribution, small-world, scale-freeness, resistance to random errors and vulnerability to targeted attacks. In the first part of this lecture, we will present a critical review of the topological properties of biochemical networks, showing that the most popular publications in network biology are based on myths rather than facts [1].

The claim that the degree (number of links) follows a power law distribution relies on a rough inspection of binned plots, but fades out as soon as the data sets are displayed in full detail, and is contradicted by statistical fitting tests. The small-world property was established by measuring the average length of the shortest paths between any pair of compounds in metabolic networks. However, the algorithms used to measure path lengths return irrelevant paths, where pool metabolites (H_2O, H^+, O_2, etc.) are used as intermediate links between reactions. The small-world measure is thus an artefact coming from the application of general graph statistics that are inadequate for metabolic networks. The properties of tolerance to random deletions and vulnerability to targeted attacks were postulated by analogy with computer networks (Internet). Tolerance of metabolic networks to random errors is however contradicted by the hundreds of auxotrophic mutants that were generated during 50 years, and allowed biochemists to isolate most of the currently known enzyme-coding genes. The concept of vulnerability to targeted attacks is simply devoid of sense as soon as one considers that removing a single "hub" (e.g. H_2O) from any organism would require deleting several hundreds of enzyme-coding genes, an event that can be realized neither in nature, nor even in laboratory-controlled conditions.

Despite the weaknesses of these foundations, topological analysis of metabolic networks can provide insight into metabolism and its evolution, but this requires the development of dedicated methods, taking into account the biochemical properties of reactions and compounds. In the second part of the

lecture, we will present various graph-based algorithms that were designed to discover relevant pathways in metabolic networks: filtering out of pool metabolites [2], weighting of compounds according to their degree [3], decomposition of reactions into reactant pairs [4], multiple-end sub-network extraction [5]. We will show a quantitative evaluation of the respective accuracy of those methods, and illustrate their practical value on the basis of selected study cases.

References

[1] Lima-Mendez, G. and van Helden, J. (2009). The powerful law of the power law and other myths in network biology. *Mol Biosyst* **5**: 1482-93.

[2] van Helden, J., Wernisch, L., Gilbert, D. and Wodak, S. J. (2002). Graph-based analysis of metabolic networks In al., M. H.-W. e. (Ed), Ernst Schering Res Found Workshop, Springer-Verlag, pp. 245-74.

[3] Croes, D., Couche, F., Wodak, S. J. and van Helden, J. (2006). Inferring meaningful pathways in weighted metabolic networks. *J Mol Biol* **356**: 222-36.

[4] Faust, K., Croes, D. and van Helden, J. (2009). Metabolic pathfinding using RPAIR annotation. *J Mol Biol* **388**: 390-414.

[5] Faust, K., Dupont, P., Callut, J. and van Helden, J. (2010). Pathway discovery in metabolic networks by subgraph extraction. *in revision*.

Modeling intracellular cargo transport by several molecular motors

Melanie Mueller[1]

[1] MPIKG Postdam, Germany

Abstract

The complex internal structure of cells depends to a large extend on active transport by molecular motors. These molecular motors are 'nano-trucks' that transport various cargoes, like vesicles, organelles or mRNA, along cytoskeletal filaments, the 'roads'.

Many cargoes are transported by small teams of about 1-10 motors. Some cargoes make use of just one team of motors of the same kind, while other cargoes are propelled by two different motor teams. These teams might move into opposite directions on the same filament, or move on different types of filaments.

In this talk, we will describe systematic stochastic models for cargo transport by one or two small teams of molecular motors. These models are based on single motor properties as determined in single molecule experiments, and can be used to explain and predict various properties of the movements of cargoes inside of cells. By providing a direct connection between the behavior of single motors and intracellular transport, the models lead to an improved understanding of this transport and its biological functions.

Physical and Mechanical Inputs into Chromosomal Processes

Nancy Kleckner[1]

[1] Harvard University, Department of Molecular and Cellular Biology
Cambridge, MA. USA

Abstract

Our laboratory is interested to understand chromosomal processes through the lens of physics and engineering. Most especially, we suppose that the phyiscal and mechanical properties of chromosomes are of central functional significance, i.e. that they play governing roles in fundamental processes. We came to this problem by considering spatial patterning of crossovers along meiotic chromosomes, which exhibits features explainable by a stress/stress relief/stress redistribution mechanism. Additional considerations led to the proposition that chromatin expansion generates inter-chromosomal "pushing forces". Recent work to be discussed involves several approaches:

i We have developed a magnetic micropiston system in which we can analyze chromatin expansion and the effects such expansion under spatial confinement (J. Fisher, unpublished).

ii Molecular dynamics analysis of the HEAT repeat scaffold of protein phosphatase PP2A, and of the full heterotrimeric PP2A enzyme, points to the possibility that catalytic activity can be governed by externally-imposed force (Grinthal et al., 2010).

iii High resolution visualization of the E.coli nucleoid in 3D in living cells over time in the cell cycle is studied in combination with precision definition of individual loci and complexes with respect to the nucleoid. Mechanical properties of the nucleoid (which is stiff and springy) are revealed along with other properties, notably porosity. These findings lead to a model for development of structure by repulsion between negatively supercoiled plectonemes under confinement. Potential solutions to other problems of chromosome dynamics, in E.coli and in eukaryotic cells emerge (A. Bourniquel and Z. Liang, unpublished). Project (2) is carried out in collaboration with Prof. Martin Karplus, Harvard University Department of Chemistry and Chemical Biology. All of these projects, and others, including analysis of the mechanism of direct DNA/DNA pairing (Danilowicz et al., 2009) and its in vivo roles, are carried out in collaboration with Prof. Mara Prentiss, Harvard University Department of Physics.

iv Another problem of current interest is the way in which, during meiosis, homologous chromosomes recognize one another and come together in space, with concomitantly generating entanglements (Koszul et al., 2008; Koszul and Kleckner, 2009; Storlazzi et al., 2010).

These and other projects are carried out in collaboration with Prof. Denise Zickler, U. Paris-Sud, Orsay, France and Dr. Aurora Storlazzi, IGB, Napoli, Italia.

To Be Announced

Eshel Ben Jacob[1]

[1] Tel-Aviv Univ., Israel

Abstract

PART II ARTICLES

Algorithmic Aspects of Analysis and Control of Boolean Networks

Tatsuya Akutsu[1]

[1] Bioinformatics Center, Institute for Chemical Research, Kyoto University

Abstract

The Boolean network is known as a discrete model of genetic networks. In this article, we focus on detection/enumeration of attractors and computation of control actions for Boolean networks. We give a brief introduction of these problems and review algorithmic results on these problems with focusing on works by the author and collaborators. For detection of attractors, we review SAT-based algorithms and simple recursive algorithms. For control of Boolean networks, we review dynamic programming algorithms for general BNs and tree-structured BNs. We also review a general approach for both problems that are based on integer linear programming.

1 Introduction

Analysis of biological information networks is an important topic in bioinformatics, computational biology, and systems biology. In order to analyze these networks, various kinds of mathematical models have been proposed. Among them, the Boolean network (BN, in short) has received much attention [15]. BN is a very simple model of genetic networks: each node corresponds to a gene and takes either 0 (inactive) or 1 (active), and the states of nodes change synchronously according to regulation rules given as Boolean functions. Thus, it is easily seen that a BN with n nodes has a total of 2^n possible global states. Since each global state deterministically transits to the same or another global state, beginning from any initial global state, the system will eventually evolve into a limited set of stable states called *attractors*. An attractor consisting of only one state is called a *singleton attractor* or a *fixed point*. Otherwise, it is called a *cyclic attractor*.

Since it is considered that attractors correspond to distinct cell states, extensive studies have been done on the distribution of attractors [9, 15, 22]. However, no conclusive results have not yet been obtained. From a computational viewpoint, not so much attention had been paid for detection and/or enumeration of attractors. However, due to the need for analyzing real genetic networks, extensive studies have recently been done on detecting and enumerating attractors.

Akutsu et al. showed that deciding existence of a singleton attractor is NP-complete and counting the number of singleton attractors is #P-complete [1]. Tošić showed that the counting problem remains #P-complete even if graphs are restricted to be planer bipartite graphs [27]. Kosub showed that the existence problem can be solved in polynomial time for several special cases relating with bounded treewidth [16]. Several heuristic methods have also been proposed for enumeration of fixed points and/or cyclic attractors [8, 10, 12, 19].

In addition to attractor problems, control problems for Boolean networks are becoming important because development of control theory/methods for biological networks is a one of the major goals of systems biology and has potential applications of systems-based drug discovery and cancer treatment [14]. Datta et al. proposed a method for finding a control strategy for Probabilistic Boolean Networks (PBNs, in short) [7] from which many extensions and variants followed [21], where PBN is a probabilistic extension of BN. In their approach, it is assumed that states of some nodes can be externally controlled and the objective is to find a sequence of control actions with the minimum cost that leads to a desirable global state. Their approach is based on the theory of Markov chains and makes use of the classical technique of dynamic programming. However, it is required in their methods to handle exponential size matrices and thus their methods can only be applied to small biological systems. Therefore, it is reasonable to ask how difficult it is to find control strategies for BNs. Akutsu et al. showed that finding control strategies for BNs (and PBNs) is NP-hard [2]. On the other hand, they showed that this problem can be solved in polynomial time if BN has a tree structure. Cheng and Qi proposed control models and methods of BNs using the concept of semi-tensor product [5]. Langmead and Jha developed a practical method based on model checking and successfully applied the method to finding of control policies to an existing model of fruit fly embryo development [18]. Recently, Akutsu et al. developed practical integer linear programming-based methods that can be applied to for both control and attractor detection problems [4]

In this article, we review algorithmic results on attractor detection and control problems on BNs, with focusing on works by the author and collaborators. In Section 2, we introduce BN and give problem definitions. Next, we review algorithms for the attractor detection problem and the control problem in Sections 3 and 4, respectively. Then, we review integer linear programming methods for both problems. Finally, we conclude with future directions.

2 Preliminaries

2.1 Boolean Network

A BN is represented by a set of *nodes* and a set of regulation rules for nodes, where each node corresponds to a gene. Each node takes either 0 or 1 at each discrete time t, where 1 (resp. 0) means that the corresponding gene is active (resp. inactive) at time t. A regulation rule for each node is given in the form of a Boolean function and the states of nodes change synchronously. Formally, a BN $G(V, F)$ consists of a set $V = \{v_1, \ldots, v_n\}$ of nodes and a list $F = (f_1, \ldots, f_n)$ of *Boolean functions*, where a Boolean function $f_i(v_{i_1}, \ldots, v_{i_k})$ with inputs from specified nodes $v_{i_1}, \ldots, v_{i_k}$ is assigned to each node v_i. We use $IN(v_i)$ to denote the set of input nodes $v_{i_1}, \ldots, v_{i_k}$ to v_i. The state of node v_i at time $t + 1$ is determined by

$$v_i(t + 1) = f_i(v_{i_1}(t), \ldots, v_{i_{k_i}}(t)).$$

We let $\mathbf{v}(t) = [v_1(t), \ldots, v_n(t)]$, which is called a *global state* or a *Gene Activity Profile* (GAP) at time t. We also write $v_i(t + 1) = f_i(\mathbf{v}(t))$ and $\mathbf{v}(t + 1) = \mathbf{f}(\mathbf{v}(t))$ to denote the regulation rules for v_i and the whole BN, respectively. We define the set of edges E by $E = \{(v_{i_j}, v_i)|v_{i_j} \in IN(v_i)\}$, and then $G(V, E)$ is a directed graph representing the network topology of a BN. The number of input nodes to v_i is called the *indegree* of v_i. We use K to denote the *maximum indegree* of a BN, which strongly affects the computational complexities in many algorithms.

An example of BN is given in Fig. 1. In this example, the state of node v_1 at time $t + 1$ is determined by the state of node v_3 at time t. The states of node v_2 and v_3 at time $t + 1$ are determined by logical AND of the state of node v_1 and negation of the state of node v_3 at time t and by logical AND of the state of node v_1 and negation of the state of node v_2 at time t, respectively. We use $x \wedge y$, $x \vee y$, $x \oplus y$, $\overline{x}$ to denote logical AND of x and y, logical OR of x and y, exclusive OR of x and y, and logical NOT of x, respectively. Dynamics of a BN is well-described by a *state transition table* and a *state transition diagram* shown in Fig. 1. For example, the second row of the table means that if the state of BN is $[0, 0, 1]$ at time t then the state will be $[1, 0, 0]$ at time $t + 1$, which is also represented by an arc from 001 to 100 in the diagram.

2.2 Attractor

Starting from an initial global state $\mathbf{v}(0)$, a BN will eventually reach a set of global states, called an *attractor*, which forms a directed cycle in the state transition diagram. An attractor consisting of only one global state (i.e., $\mathbf{v} = \mathbf{f}(\mathbf{v})$) is called a *singleton attractor* or a *fixed point*. Otherwise, it is called a *cyclic*

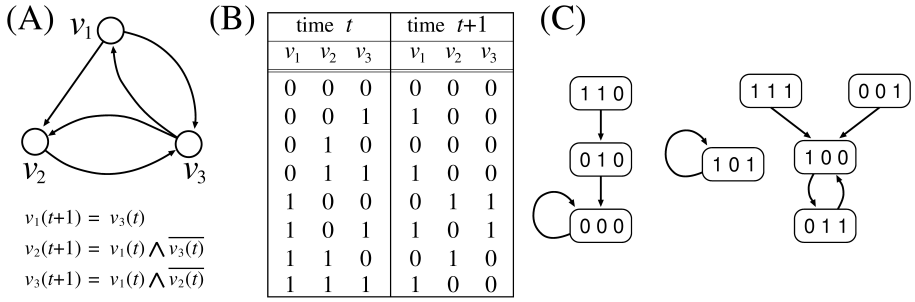


Figure 1: Example of a Boolean network. Dynamics of BN (A) is well-described by a state transition table (B) and by a state transition diagram (C).

attractor with period p if it consists of p global states $\{\mathbf{v}^1, \mathbf{v}^2, \ldots, \mathbf{v}^p\}$. For example, in Fig. 1, 000 and 101 are singleton attractors, whereas $\{011, 100\}$ is a cyclic attractor with period 2.

2.3 Problem Definitions

Based on the above definition of attractors, the attractor detection problem is defined as follow.

Definition 1 [Attractor Detection]
Instance: *a BN and the maximum length of period p_{max},*
Problem: *find an attractor with period at most p_{max}. If there does not exist such an attractor, "None" should be output.*

In this article, we mainly consider the case of $p = 1$ because it is most fundamental and there is no good way to cope with the cases of non-small p.

Akutsu et al. introduced the problem of control of BN [2], by specializing the control problem for PBN [7]. In control of BN, nodes are divided into two types, *internal nodes* and *external nodes*, where internal nodes correspond to usual nodes in a BN and external nodes correspond to control nodes for which we can arbitrarily specify the states at any time step. Let $V = \{v_1, \ldots, v_n, v_{n+1}, \ldots, v_{n+m}\}$, where $v_1, \ldots, v_n$ are internal nodes and $v_{n+1}, \ldots, v_{n+m}$ are external nodes. For convenience, we use u_i to denote an external node v_{n+i}. Then, states of internal nodes ($v_i(t+1)$ for $i = 1, \ldots, n$) are determined by

$$v_i(t+1) = f_i(v_{i_1}(t), \ldots, v_{i_{k_i}}(t)),$$

where each v_{i_k} is either an internal node or an external node. Here, we let $\mathbf{v}(t) = [v_1(t), \ldots, v_n(t)]$ and $\mathbf{u}(t) = [u_1(t), \ldots, u_m(t)]$. We can describe the state transition rule of a BN by

$$\mathbf{v}(t+1) = \mathbf{f}(\mathbf{v}(t), \mathbf{u}(t)),$$

where $\mathbf{u}(t)$s are determined externally. Then, the control problem is defined as follows (see also Fig. 2).

Definition 2 [Control of BN]
Instance: *a BN, an initial state of the network for internal nodes* $\mathbf{v}^0$*, and the desired state of the network for internal nodes* $\mathbf{v}^M$ *at the* M*-th time step,*
Problem: *find a sequence of 0-1 vectors* $\langle \mathbf{u}(0), \ldots, \mathbf{u}(M) \rangle$ *such that* $\mathbf{v}(0) = \mathbf{v}^0$ *and* $\mathbf{v}(M) = \mathbf{v}^M$*. If there does not exist such a sequence, "None" should be the output.*

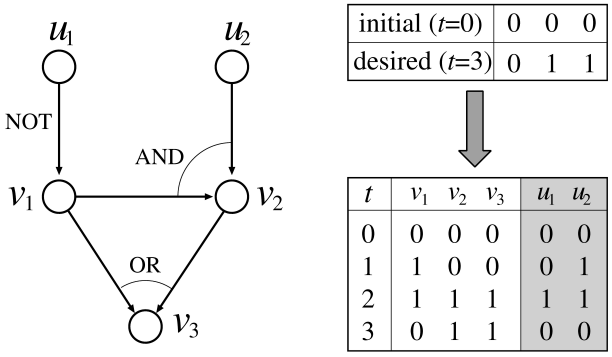


Figure 2: Example of control of a Boolean network. In this problem, given initial and desired states of internal nodes (v_1, v_2, v_3), it is required to compute a sequence of states of external nodes (u_1, u_2) leading to the desired state.

3 Algorithms for Attractor Detection

Since there are 2^n global states for a BN of n nodes, the attractor detection and enumeration problems can be solved in $O(2^n poly(n))$ time (under the assumption that the value of each Boolean function can be calculated in a polynomial time) by constructing a transition diagram. Therefore, we are interested in developing $o(2^n)$ time algorithms for attractor problems. Though it seems quite difficult to develop such an algorithm for the general case, it is possible in some reasonably restricted cases. In this section, we review such algorithms.

3.1 Simple Reduction to SAT

Since BN is based on predicate logic, it is reasonable to try to apply existing algorithms developed for predicate logic to attractor detection and enumeration problems. In particular, it would be useful to apply algorithms for SAT (Boolean satisfiability problem) because extensive studies have been done and a number of algorithms have been developed for solving SAT [6, 13, 17, 28].

Based on the observation in [19], Tamura and Akutsu showed that the singleton attractor detection problem for BNs with maximum indegree K can be transformed into $(K+1)$-SAT with n variables [24] in a simple manner. k-SAT is a well studied problem in theoretical computer science and is defined as: given a set of clauses (i.e., a set of disjunctions of literals) over a set of Boolean variables, decide whether or not there exists a 0-1 assignment to variables that satisfies all the clauses, where each clause consists of at most k literals.

Here, we only show a reduction procedure for the cases of BNs with maximum indegree 2. Let $\mathbf{v}$ be a global state of a BN. Recall that $\mathbf{v}$ is a singleton attractor if $v_i = f_i(\mathbf{v})$ holds for all $i = 1, \ldots, n$. Here, we also use v_i to denote the state of v_i because $v_i(t) = v_i(t+1) = v_i(t+2) = \cdots$ holds in a singleton attractor. In the following, l_i denotes either v_i or $\overline{v_i}$. We begin with the empty set. For $i = 1$ to n, we add clause(s) to the set according to the following rules (we omit the cases of constant and unary functions)

$$
\begin{aligned}
v_i = l_j \vee l_k &\iff (\overline{v_i} \vee l_j \vee l_k) \ \wedge \ (v_i \vee \overline{l_j \vee l_k}) \\
&\iff (\overline{v_i} \vee l_j \vee l_k) \ \wedge \ (v_i \vee (\overline{l_j} \wedge \overline{l_k})) \\
&\iff (\overline{v_i} \vee l_j \vee l_k) \ \wedge \ (v_i \vee \overline{l_j}) \ \wedge \ (v_i \vee \overline{l_k}), \\
v_i = l_j \wedge l_k &\iff (\overline{v_i} \vee (l_j \wedge l_k)) \ \wedge \ (v_i \vee (\overline{l_j \wedge l_k})) \\
&\iff (\overline{v_i} \vee l_j) \ \wedge \ (\overline{v_i} \vee l_k) \ \wedge \ (v_i \vee \overline{l_j} \vee \overline{l_k}), \\
v_i = l_j \oplus l_k &\iff (\overline{v_i} \vee ((l_j \vee l_k) \wedge (\overline{l_j} \vee \overline{l_k})) \ \wedge \ (v_i \vee \overline{(l_j \vee l_k) \wedge (\overline{l_j} \vee \overline{l_k})}) \\
&\iff (\overline{v_i} \vee l_j \vee l_k) \ \wedge \ (\overline{v_i} \vee \overline{l_j} \vee \overline{l_k}) \\
&\qquad \wedge (v_i \vee (\overline{l_j \vee l_k}) \vee (\overline{\overline{l_j} \vee \overline{l_k}})) \\
&\iff (\overline{v_i} \vee l_j \vee l_k) \ \wedge \ (\overline{v_i} \vee \overline{l_j} \vee \overline{l_k}) \\
&\qquad \wedge (v_i \vee \overline{l_j} \vee l_k) \ \wedge \ (v_i \vee l_j \vee \overline{l_k}).
\end{aligned}
$$

Then, we can see that a regulation rule for a node v_i is transformed into at most four clauses in 3-SAT. Therefore, the singleton attractor detection problem for BNs with maximum indegree 2 is reduced to 3-SAT with n variables and at most $4n$ clauses. This transformation can be generalized to arbitrarily fixed K.

Proposition 1 [24] *Any instance of the singleton attractor detection problem for a BN of maximum indgree K with n nodes can be reduced in polynomial time to an instance of $(K+1)$-SAT with at most $2^{K+1} \cdot n$ clauses and n variables.*

This result can be extended for the cyclic attractor detection problem with period p by encoding $\mathbf{f}^p(\mathbf{v})$ using (K^p+1)-SAT clauses.

Theorem 1 [3] *Any instance of the cyclic attractor detection problem with period p for a BN of maximum indgree K with n nodes can be reduced in polynomial time to an instance of (K^p+1)-SAT with at most $(\sum_{p=1}^{K} 2^{K^p+1}) \cdot n$ clauses and n variables.*

Combining this result with $o(2^n)$ time algorithms for k-SAT [6], we can see that the attractor detection problem can be solved in $o(2^n)$ time for fixed p and K.

3.2 Algorithms for AND/OR BNs

In the above, we considered BNs with bounded maximum indegree. However, SAT algorithms can be used for developing algorithms for other special cases of BNs. Indeed, $o(2^n)$ time algorithms were developed for AND/OR BNs [24, 25], in which each Boolean function is limited to AND or OR of literals whereas there is no restriction on the maximum indgree [24, 25]. Here, we briefly review the basic idea used in these algorithms.

Suppose that the following Boolean function is assigned to a node v_i:

$$v_i(t+1) = v_1(t) \wedge v_2(t) \wedge \cdots \wedge v_h(t).$$

Among four possible assignments $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ for (v_1, v_i), three satisfy the condition of a singleton attractor whereas one (i.e., $(0,1)$) does not. Therefore, we can eliminate two nodes by examining these three assignments. If we could continue this procedure until there is no remaining node, the complexity of $O(3^{(n/2)}) \approx O(1.733^n)$ would be achieved by solving

$$g(2) = 3, \quad g(k) = 3 \cdot g(k-2).$$

However, we cannot continue the above mentioned procedure if there is no remaining edge and only singleton nodes are left. In order to cope with such a case, the following algorithm was developed [24] by utilizing an algorithm for SAT with m clauses [28].

1. Let all the nodes be non-assigned.
2. While there exists a non-assigned node pair $(u, v) \in E$, examine all possible 3 assignments on $(u.v)$ recursively.
3. Let U be the set of nodes whose values were already assigned.
4. If $|U| > \alpha n$, examine all possible assignments on the remaining nodes and then check the condition of a singleton attractor. Otherwise, compute an appropriate assignment using [28] and then check the condition of a singleton attractor.

By letting $\alpha = 0.767$, it is shown that this algorithm works in $O(1.792^n)$ time [24]. This result is improved as below.

Theorem 2 [25] *The singleton attractor detection problem for AND/OR BNs can be solved in* $O(1.757^n)$.

3.3 Simple Recursive Algorithms

Though SAT-based algorithms might be useful for attractor detection problems, these might not be so useful for attractor enumeration problems. Therefore, several algorithms have been developed [29] for enumerating singleton attractors and cyclic attractors with short periods, which do not use SAT algorithms. In this subsection, we briefly review a basic version (called *basic recursive algorithm*) of these algorithms.

The number of singleton attractors in a BN depends on the regulatory rules of the BN. If the rules are $v_i(t+1) = v_i(t)$ for all i, the number of singleton attractor is 2^n. Thus, it would take at least $O(2^n)$ time in the worst case if we need to enumerate all the singleton attractors. On the other hand, it is known that the average number of singleton attractors is 1 regardless of n and K [11]. The basic recursive algorithm was designed based on these facts. It examines much smaller number of global states than 2^n on the average.

In the algorithm, a partial global state (i.e., $[v_1, \ldots, v_m]$ for $m < n$) is extended one by one towards a complete global state (i.e., singleton attractor), according to a given ordering of nodes (i.e., a random ordering). As soon as it is found that a partial global state cannot be extended to a singleton attractor, the next partial global state is examined. The pseudocode of the algorithm [29] is given below, where it is invoked with $m = 1$.

Procedure *EnumerateSingletonAttractor*(v, m)
  if $m = n + 1$ **then** Output $[v_1, v_2, \cdots, v_n]$ and **return**;
  for $b = 0$ **to** 1 **do**
    $v_m := b$;
    if it is found that $f_i(\mathbf{v}) \neq v_i$ for some $i \leq m$ **then continue**
    else *EnumerateSingletonAttractor*$(v, m + 1)$;
  return

Table 1: Average case time complexities of basic, outdegree-based, and BFS-based algorithms for singleton attractor detection [29].

K	2	3	4	5	6	7	8
basic	1.35^n	1.43^n	1.49^n	1.53^n	1.57^n	1.60^n	1.62^n
outdegree-based	1.19^n	1.27^n	1.34^n	1.41^n	1.45^n	1.48^n	1.51^n
BFS-based	1.16^n	1.27^n	1.35^n	1.41^n	1.45^n	1.50^n	1.53^n

Here we briefly analyze the average case time complexity. Assume that we have tested the first m nodes, where $m \geq K$. For all $i \leq m$, $f_i(\mathbf{v}) \neq v_i$ holds with probability

$$P(f_i(\mathbf{v}) \neq v_i) \;=\; 0.5 \cdot \frac{\left(\begin{array}{c} m \\ k_i \end{array} \right)}{\left(\begin{array}{c} n \\ k_i \end{array} \right)} \;\approx\; 0.5 \cdot (\frac{m}{n})^{|IN(v_i)|} \;\geq\; 0.5 \cdot (\frac{m}{n})^K,$$

where we assume that Boolean functions of $|IN(v_i)|$ ($\leq K$) inputs are selected at uniformly random. If $f_i(\mathbf{v}) \neq v_i$ holds for some $i \leq m$, the algorithm cannot proceed to the next recursive level. Therefore, the probability that the algorithm examines the $(m+1)$-th node is no more than

$$[1 - P(f_i(\mathbf{v}) \neq v_i)]^m \;=\; [1 - 0.5 \cdot (\frac{m}{n})^K]^m.$$

By means of numerical calculation for estimating the maximum of this expression, the average case time complexities are estimated for $K = 2, \ldots, 7$ as in the first row of Table 1.

Several variants of this basic recursive algorithm are proposed in [29], by changing the order of sorting nodes before invoking the recursive procedure. For the orderings based on the outdegree and BFS (breadth-first search), theoretical estimates of the average case time complexity are obtained as in Table 1. Computational experiments were performed in order to verify these theoretical results, and good agreements were observed [29]. The basic recursive algorithm was extended for enumeration of cyclic attractors with short periods [29].

4 Algorithms for Control of Boolean Networks

Datta et al. proposed a dynamic programming based method for finding a control strategy for PBN [7]. Since their method can also be applied to BN. we briefly review it in the context of BN.

We use a table $D[b_1, \ldots, b_n, t]$, where each entry takes either 0 or 1. $D[b_1, \ldots, b_n, t]$ takes 1 if there exists a control sequence $\langle \mathbf{x}(t), \ldots, \mathbf{x}(M) \rangle$ leading to the target state $\mathbf{v}^M$ beginning from the state $[b_1, \ldots, b_n]$ at time t. This table is computed from $t = M$ to $t = 0$ by using the following dynamic programming procedure:

$$D[b_1, \ldots, b_n, M] = \begin{cases} 1, & \text{if } [b_1, \ldots, b_n] = \mathbf{v}^M, \\ 0, & \text{otherwise,} \end{cases}$$

$$D[b_1, \ldots, b_n, t-1] = \begin{cases} 1, & \text{if there exists } (\mathbf{c}, \mathbf{u}) \text{ such that} \\ & D[c_1, \ldots, c_n, t] = 1 \text{ and } \mathbf{c} = \mathbf{f}(\mathbf{b}, \mathbf{u}), \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{b} = [b_1, \ldots, b_n]$ and $\mathbf{c} = [c_1, \ldots, c_n]$. Then, there exists a desired control sequence if and only if $D[a_1, \ldots, a_n, 0] = 1$ holds for $\mathbf{v}^0 = [a_1, \ldots, a_n]$. Once this table is constructed, a desired control sequence can be obtained using the standard *traceback* technique.

Now, we estimate the time complexity. The size of table $D[b_1, \ldots, b_n, t]$ is clearly $O(M \cdot 2^n)$. Since we should examine pairs of $O(2^n)$ internal states and $O(2^m)$ external states for each time t, it requires $O(M \cdot 2^{n+m})$ time excluding the time for calculation of Boolean functions. This time complexity is not practical even for medium size BNs (e.g., $n + m \geq 30$).

Since control of BN is NP-hard [1], exponential time is almost inevitable for the general case. However, it may be possible to develop polynomial time algorithms for special cases. Akutsu et al. developed such an algorithm for the case where the network has a tree structure (i.e., the graph is connected and there is no cycle) [2]. Since that algorithm is a bit involved, we review here a simplified algorithm for the case where the network has a rooted tree structure (i.e., all paths are directed from leaves to the root).

In order to compute a control strategy, we employ dynamic programming in a different way than in [7]. We define a table $S[v_i, t, b]$ as below, where v_i is a node in a BN, t is a time step and b is a Boolean value (i.e., 0 or 1). Here $S[v_i, t, b]$ takes 1 if there exists a control sequence (up to time t) that makes $v_i(t) = b$.

$$S[v_i, t, 1] = \begin{cases} 1, & \text{if there exists } \langle \mathbf{x}(0), \ldots, \mathbf{x}(t) \rangle \text{ such that } v_i(t) = 1, \\ 0, & \text{otherwise.} \end{cases}$$

$$S[v_i, t, 0] = \begin{cases} 1, & \text{if there exists } \langle \mathbf{x}(0), \ldots, \mathbf{x}(t) \rangle \text{ such that } v_i(t) = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then, $S[v_i, t, 1]$ can be calculated by the following dynamic programming procedure.

$$S[v_i, t+1, 1] = \begin{cases} 1, & \text{if there exists } [b_{i_1}, \ldots, b_{i_k}] \text{ such that } f_i(b_{i_1}, \ldots, b_{i_k}) = 1 \\ & \text{holds and } S[v_{i_j}, t, b_{i_j}] = 1 \text{ holds for all } j = 1, \ldots, k, \\ 0, & \text{otherwise.} \end{cases}$$

$S[v_i, t, 0]$ can be calculated in a similar way. It is to be noted that each leaf is either a constant node or an external node. For a constant node, either $S[v_i, t, 1] = 1$ and $S[v_i, t, 0] = 0$ hold for all t, or $S[v_i, t, 1] = 0$ and $S[v_i, t, 0] = 1$ hold for all t. For an external node, $S[v_i, t, 1] = 1$ and $S[v_i, t, 0] = 1$ hold for all t. Since the size of table $S[v_i, t, b]$ is $O((n+m)M)$, this dynamic programming algorithm works in polynomial time where we assume that the value of each Boolean function can be computed in polynomial time. A desired control sequence can also be obtained from the table in polynomial time using the standard traceback technique. This algorithm was extended for BNs with general tree structures.

Theorem 3 [2] *Control of BN can be solved in polynomial time if BN has a tree structure.*

5 Integer Linear Programming-Based Approach

We have discussed so far theoretical approaches to attractor detection and control of BN. However, practical methods should also be developed. In this section, we briefly review a general approach based on *integer linear programming* (ILP) [4]. ILP is to optimize (maximize or minimize) a linear function under a set of linear constraints (linear inequalities) and the condition that specified variables must take integer values.

Here, we review the key idea of the approach using a simple example. Suppose that the regulation rule for a node v_3 is given as $v_3(t+1) = f_3(v_1(t), v_2(t)) = v_1(t) \oplus v_2(t)$. Let v_i' and v_i be 0-1 variables denoting the states of $v_i(t+1)$ and $v_i(t)$, respectively. Then, this regulation rule is transformed into the following Boolean formula

$$\begin{aligned} f_3(v_1, v_2) &= (f_3(0,0) \wedge \overline{v_1} \wedge \overline{v_2}) \vee (f_3(0,1) \wedge \overline{v_1} \wedge v_2) \vee (f_3(1,0) \wedge v_1 \wedge \overline{v_2}) \\ &\quad \vee (f_3(1,1) \wedge v_1 \wedge v_2) \\ &= (\overline{v_1} \wedge v_2) \vee (v_1 \wedge \overline{v_2}). \end{aligned}$$

This Boolean formula is further transformed into the following inequalities

$$\begin{aligned} v_{3,00} &= 0, \\ v_{3,01} &\geq (1 - v_1) + v_2 - 1 = v_2 - v_1, \end{aligned}$$

$$
\begin{aligned}
v_{3,01} &\leq \frac{1}{2}(1 - v_1 + v_2), \\
v_{3,10} &\geq v_1 + (1 - v_2) - 1 \;=\; v_1 - v_2, \\
v_{3,10} &\leq \frac{1}{2}(v_1 + 1 - v_2), \\
v_{3,11} &= 0, \\
v'_3 &\leq v_{3,00} + v_{3,01} + v_{3,10} + v_{3,11}, \\
v'_3 &\geq \frac{1}{4}(v_{3,00} + v_{3,01} + v_{3,10} + v_{3,11}),
\end{aligned}
$$

where all of $v_{3,ij}$ are 0-1 integer variables. In the case of single attractor detection, we let $v'_i = v_i$ because $\mathbf{v}(t) = \mathbf{v}(0)$ holds for all t.

The results of computational experiments suggest for both singleton attractor detection and control of BN that this ILP-based approach can be applied to large-size BNs if $K \leq 2$ whereas it can only be applied to BNs with several tens of nodes if $K = 3$ [4].

6 Concluding Remarks

We have reviewed algorithms developed for attractor detection and control of Boolean networks with focusing on works done by the author and collaborators. For attractor detection, we reviewed $o(2^n)$ time algorithms. However, these algorithms are not necessarily optimal and it seems that there exists much room for improvements. Therefore, improvement of these algorithms is left as an open problem. In particular, development of much more efficient algorithms for cyclic attractor detection is left as an important open problem. For control of BN, there exists only a few complexity results (especially, a very few positive results). Therefore, development of polynomial time or $o(2^n)$ time algorithms for wider classes of BNs is left as an open problem.

In this article, BNs are regarded as a model of genetic networks. However, BN and its variants can also be used as models of other types of biological networks. Recently, BN-like models were proposed for modeling metabolic networks [20, 23, 26], in which chemical compounds and chemical reactions are regarded as OR nodes and AND nodes, respectively. Using these models, problems of deciding the minimum number of chemical reactions/enzymes to be inactivated for preventing production of specified chemical compounds were studied, which may have potential applications to identification of multiple drug targets. ILP-based methods were also developed for these problems [20, 26].

There is a criticism that BN is too simple as a model of genetic networks, metabolic networks, and/or other types of biological networks. However, studies on BNs may provide some insights into other models. At least, hardness

results should hold for more general models. Some ideas in theoretical and/or practical algorithms for BNs might also be useful for design and analysis of algorithms for more general models. Therefore, extension of BNs along with efficient practical and/or theoretical algorithms is an important future direction.

References

[1] T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano, A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions, *Genome Informatics*, 9:151–160, 1998.

[2] T. Akutsu, M. Hayashida, W-K. Ching, and M. K. Ng, Control of Boolean networks: hardness results and algorithms for tree-structured networks, *Journal of Theoretical Biology*, 244:670–679, 2007.

[3] T. Akutsu, Detection and enumeration of steady states in biological information networks, *RIMS Kokyuroku*, 1644:110-119, 2009.

[4] T. Akutsu, M. Hayashida, and T. Tamura, Integer programming-based methods for attractor detection and control of Boolean networks, *Proc. The combined 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference*, in press.

[5] D. Cheng and H. Qi, Controllability and observability of Boolean control networks, *Automatica*, 45:1659-1667, 2009.

[6] E. Dantsin, A. Goerdt, E. A. Hirsch, R. Kannan, J. M. Kleinberg, C. H. Papadimitriou, P. Raghavan, and U. Schöning, A deterministic $(2-2/(k+1))n$ algorithm for k-SAT based on local search, *Theoretical Computer Science*, 289:69–83, 2002.

[7] A. Datta, A. Choudhary, M. L. Bittner, and E. R. Dougherty, External control in Markovian genetic regulatory networks, *Machine Learning*, 52:169–191, 2003.

[8] V. Devloo, P. Hansen, and M. Labbé, Identification of all steady states in large networks by logical analysis, *Bulletin of Mathematical Biology*, 65:1025–1051, 2003.

[9] B. Drossel, T. Mihaljev, and F. Greil, Number and length of attractors in a critical Kauffman model with connectivity one, *Physical Review Letters*, 94:088701, 2005.

[10] A. Garg, A. DiCara, I. Xenarios, L. Mendoza, and G. DeMicheli, Synchronous versus asynchronous modeling of gene regulatory networks, *Bioinformatics*, 24:1917–1925, 2008.

[11] I. Harvey and T. Bossomaier, Time out of joint: Attractors in asynchronous random boolean networks, *Proc. 4th European Conference on Artificial Life*, 67–75, 1997.

[12] D. J. Irons, Improving the efficiency of attractor cycle identification in Boolean networks, *Physica D*, 217:7–21, 2006.

[13] K. Iwama, Worst-case upper bounds for kSAT, *Bulletin of the EATCS*, 82:61–71, 2004.

[14] H. Kitano, Cancer as a robust system: implications for anticancer therapy, *Nature Reviews Cancer*, 4:227–235, 2004.

[15] S. A. Kauffman, *The Origins of Order: Self-organization and Selection in Evolution*, Oxford Univ. Press, New York, 1993.

[16] S. Kosub, Dichotomy results for fixed-points existence problems for Boolean dynamical systems, *Mathematics in Computer Science*, 1:487–505, 2008.

[17] K. Kutzkov, New upper bound for the #3-SAT problem, *Information Processing Letters*, 105:1–5, 2007.

[18] C. J. Langmead and S. K. Jha, Symbolic approaches for finding control strategies in Boolean networks, *Journal of Bioinformatics and Computational Biology*, 7:323–338, 2009.

[19] M. Leone, A. Pagnani, G. Parisi, and O. Zagordi, Finite size corrections to random Boolean networks, *Journal of Statistical Mechanics: Theory and Experiment*, P12012, 2006.

[20] Z. Li, R.-S. Wang, X.-S. Zhang, and L. Chen, Detecting drug targets with minimum side effects in metabolic networks, *IET Systems Biology*, 3:523–533, 2009.

[21] R. Pal, A. Datta, and E. R. Dougherty, Robust intervention in probabilistic Boolean networks, *IEEE Transactions on Signal Processing*, 56:1280–1294, 2008.

[22] B. Samuelsson and C. Troein, Superpolynomial growth in the number of attractors in kauffman networks, *Physical Review Letters*, 90:098701, 2003.

[23] P. Sridhar, B. Song, T. Kahveci, and S. Ranka, Mining metabolic networks for optimal drug targets, *Proc. Pacific Symposium on Biocomputing*, 291–302, 2008.

[24] T. Tamura and T. Akutsu, Detecting a singleton attractor in a Boolean network utilizing SAT algorithms, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E92-A:493–501, 2009.

[25] T. Tamura and T. Akutsu, Algorithms for singleton attractor detection in planar and nonplanar AND/OR Boolean networks, *Mathematics in Computer Science*, 2:401–420, 2009.

[26] T. Tamura, K. Takemoto, and T. Akutsu, Finding minimum reaction cuts of metabolic networks under a Boolean model using integer programming and feedback vertex sets, *International Journal of Knowledge Discovery in Bioinformatics*, in press.

[27] P. Tošić, On the complexity of counting fixed points and gardens of Eden in sequential dynamical systems on planar bipartite graphs, *International Journal of Foundations of Computer Science*, 17:1179–1203, 2006.

[28] M. Yamamoto, An improved $\tilde{O}(1.234^m)$-time deterministic algorithm for SAT, *Proc. 16th International Symposium on Algorithms and Computation*, 644–653, 2005.

[29] S-Q. Zhang, M. Hayashida, T. Akutsu, W-K. Ching, and M. K. Ng, Algorithms for finding small attractors in Boolean networks, *EURASIP Journal on Bioinformatics and Systems Biology*, 2007:20180, 2007.

Introducing continuous time in discrete models of gene regulatory networks.

Jean-Paul Comet[1] and Gilles Bernot[1]

[1] I3S Laboratory, UMR 6070 CNRS and University of Nice Sophia Antipolis, 2000 route des Lucioles, B.P. 121, F-06903 Sophia Antipolis CEDEX

Abstract

It is becoming a routine task to build models of increasing complexity about a given gene network. While available data on the connectivity between elements of the network are more and more numerous, the kinetic data of the associated interactions remain difficult to interprete in order to identify the strength of the gene activations or inhibitions. This parameter identification problem constitutes the cornerstone of the modelling processes. In this article, we show that some information about the elapsed time that takes a trajectory between two points (and that can be experimentally measured) can be of great interest for constraining the parameters of the model. It brings us to set out various frameworks of hybrid modelling (where a model is defined as a combination of a qualitative model with additional continuous variables) in which it is possible to compute elapsed time of trajectories while maintaining powerful automated reasoning capacities. This chapter is an overview of the main formal frameworks able to treat activation or inhibition delays between genes.

1 Introduction

Computational modelling of gene regulatory networks aims at deep understanding of how their components are controlled, thus allowing the prediction of a set of non-obvious behaviours that can be experimentally tested. Unfortunately, while available data on the interaction graph between genes are more and more numerous, the kinetic data allowing us to identify the sensible parameter values are difficult to obtain experimentally and they require many indirect reasonings. This parameter identification problem constitutes the cornerstone of the modelling activities. More precise the available information about the dynamics of the system, more precise can be the model. But precision is not the main criterion. If the precision of the model is higher than the one of the knowledge of the biological system, the precision given by computer simulations is only a consequence of an arbitrary choice of parameter values. Qualitative models where parameters are easier to identify constitute the good compromise.

This comment motivated some researchers to develop methods where this identification problem is tractable. In particular René Thomas' discrete modelling [26] of gene regulatory networks (GRN) is a well-known approach to study the dynamics resulting from a set of interacting genes. It deals with some *discrete* parameters that reflect the possible targets of trajectories. Those parameters are *a priori* unknown, but they can generally be deduced from a well-chosen set of biologically observed trajectories.

Besides, it neglects the time delay necessary for a gene to pass from one level of expression to another one, whereas information on the time necessary for the system to go from one state to another one is often experimentally available. For example, time used by the system to cover a whole turn of a periodic trajectory (*e.g.* circadian cycle) is often available. Time can also be an abstract time such as the current state of accomplishment within a phase (*e.g.* cellular cycle where "time" is connected to the measure of the mass of the cell). Such an information is not used to face up to the parameter identification problem in the "standard" Thomas' framework without delays. Such kind of information motivates several researchers to propose formal frameworks where time is explicit.

Hybrid extensions of the discrete approach of R. Thomas make time explicit: New parameters, *i.e.* delays mandatory for a gene to go from a discrete abstract level to another one, allow the determination of time along a trajectory. Hybrid modelling frameworks preserve powerful computer-aided reasoning capabilities. Adding delays, the identification problem is more difficult because of the increased number of parameters. Nonetheless computer is able to reject a large class of parameter values.

So, hybrid models (where the levels of expression of each gene remain abstracted into a finite number of possible values but where the delays elapsed inside each discrete level are continuous real numbers) seem to be the best trade-off between *precision* and *automated reasoning* capabilities :

- Differential equations give a full continuous precision both on the concentration level of the gene products and on the time along a trajectory, but parameter values are almost impossible to identify precisely with respect to the experimental and measurement capabilities in biology, and computers are unable to perform proofs on these models, they only perform simulations.
- The discrete approach (so called "logical approach") of René Thomas provides an easy way to identify, exhaustively and using computer proofs, the sensible parameter values, but the discrete models give rise to some trajectories which cannot be observed biologically because the *in vivo* delays make them impossible.

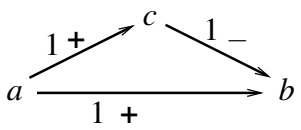


Figure 1: The incoherent type 1 feedforward loop (I1-FFL)

- *In vivo* cell models are somehow "in between" the differential equation models and the discrete models:
  - The number of molecules produced by a gene is finite and it can sometimes be very low, thus, the continuous differential models induce an abuse of precision (which can exhibit some limit behaviours that do not exist *in vivo*)
  - The number of molecules produced by a gene is most of the time much higher than the number of discrete levels in the Thomas' models, thus, discrete modelling is a rough approximation.

It appears to be possible to define adequate hybrid frameworks for the modelling of gene networks, but the task is not so easy. Many obstacles have been encountered by us and our colleagues. This chapter is an overview of the main techniques that have been proposed; it shows the main obstacles and gives a picture of the current state of the art in hybrid modelling of gene networks.

In this chapter we first present in Section 2 the basic discrete modelling framework of gene regulatory networks without delays due to R. Thomas and an extension based on formal methods from computer science which allow the automation of the search of parameter values from experimentally known behaviours. Section 3 focuses on the now classical framework of piecewise linear differential equations and their relationships with discrete models. In Section 4 we present the first hybrid approach due to R. Thomas consisting in completing a discrete model by a set of clocks which measure the time necessary to pass through a transition. Another dual approach has been also proposed by Bockmayr and Siebert [23] and is sketched in Section 5. An alternative hybrid framework is then proposed in Section 6 in which the delays introduced in the hybrid models are coherent with the underlying piecewise linear differential equation systems. Finaly we discuss in Section 7 some parameter identification issues when considering hybrid models with delays.

In order to evaluate the consequences of introducing delays into the modelling framework, we consider in the sequel some examples all based on a particular graph pattern [22]: the feedforward loop - incoherent type 1 (I1-FFL), see Figure 1, which is one of the most common network motifs. The dynamics of such a pattern of interaction graph have been largely studied [20].

The feed-forward loop is composed of a transcription factor a that regulates a second transcription factor c and both a and c regulate a gene b. So, a regulates b *via* two paths. When the signs of both paths (that is the product of signs of each interaction along the path) are not equal, the feed-forward is said incoherent [3].

Intuitivelly, it is straightforward to comprehend that when a is switched on, both b and c are subject to change. If the delay mandatory for b to come on is less than the one associated with c, then one can observe a transitory presence of b before the presence of c inhibits b. We will also see that the situation is complex when a oscillates.

2 Discrete modelling of gene regulatory networks

René Thomas has introduced in the 70's a qualitative approach [26] in order to model gene networks and to predict their dynamics. Three main ideas constitute the foundation of this qualitative approach.

2.1 First idea

The criterion to abstract the qualitative concentration levels of a gene product is the number of other genes on which it acts in the network.

Such a criterion is based on the fact that, when a gene acts on another one, the curve that represents the production rate of the target gene with respect to the concentration level of the source gene is a sigmoid. For example, assume that x is a gene that activates a gene y and inhibits a gene z as in Figure 2, then the corresponding sigmoids allow us to consider two thresholds inside the interval of all possible real concentrations levels of the x product: τ_1 and τ_2.

These two thresholds delimit three intervals within which the gene x behaves uniformly. Each interval is conventionaly identified by an integer, which is the number of genes on which x has an action. If we know the order between the thresholds, then we can additionally label the action of x on a gene by the number of the first interval that activates this action (lower left drawing of Figure 2).

So, the interaction graph contains *variables* that mostly represent genes (sometimes they represent abstract phenotypes or environmental conditions) and it contains *edges* between variables that can be labelled by a sign ($+$ for activation, $-$ for inhibition) and by an integer threshold.

2.2 Second idea

At a given global state of the network, the concentration toward which a gene product tries to go depends only on the inventory of the activations and inhibitions that act on this gene.

For example, if u is an activator of x and v is an inhibitor of x and if we assume that x has no other activator or inhibitor in the network, as in Figure 3, then four cases have to be considered:

K_x is the number of the interval toward which x tends to go when it has no help at all from the considered network. It means that u does not activate x, thus the current state of u is strictly less than the threshold of $(u \to x)$, and v inhibits x, thus the current state of v is greater or equal to the threshold of $(v \to x)$.

$K_{x,u}$ is the number of the interval toward which x tends to go when it benefits only from the help of u. It means that the current state of u is greater or equal to the threshold of $(u \to x)$, and the current state of v is greater or equal to the threshold of $(v \to x)$.

$K_{x,v}$ is the number of the interval toward which x tends to go when it benefits only from the help of v. It means that the current state of u is strictly less than the threshold of $(u \to x)$, and the current state of v is strictly less than the threshold of $(v \to x)$.

$K_{x,uv}$ is the number of the interval toward which x tends to go when it benefits both from the help of u and from the help of v. It means that the current state of u is greater or equal to the threshold of $(u \to x)$, and the current state of v is strictly less than the threshold of $(v \to x)$.

It is also possible that a gene influences itself in a given network, nevertheless auto-regulations do not change the approach at all. For each state, the parameters $K_{\dots}$ define the vector state toward which the system tends to go. Figure 4 gives a small example of gene network where we have arbitrarily chosen the parameters as follows: $K_x = 0$, $K_{x,x} = K_{x,y} = K_{x,xy} = 2$ and $K_y = K_{y,x} = 1$.

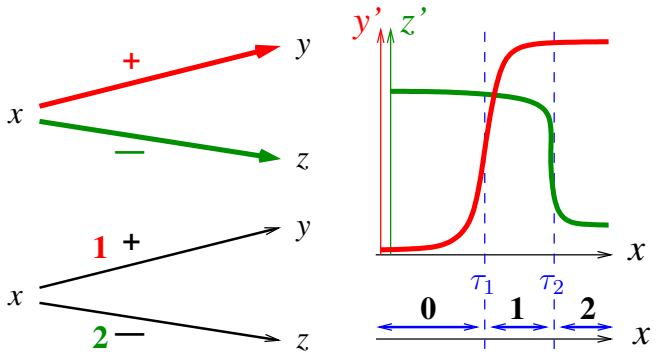
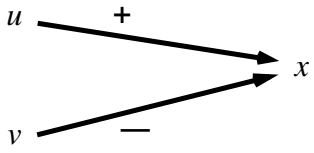
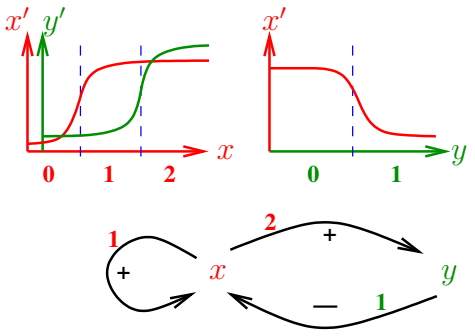


Figure 2: Multivalued regulatory graph



No help for x: K_x
The presence of u helps x: $K_{x,u}$
The absence of v helps x: $K_{x,v}$
Both help x: $K_{x,uv}$

Figure 3: Parameters



(x,y)	*Focal Point*
(0,0)	$(K_{x,y}, \quad K_y) = (2,1)$
(0,1)	$(K_x, \quad K_y) = (0,1)$
(1,0)	$(K_{x,xy}, \quad K_y) = (2,1)$
(1,1)	$(K_{x,x}, \quad K_y) = (2,1)$
(2,0)	$(K_{x,xy}, K_{y,x}) = (2,1)$
(2,1)	$(K_{x,x}, \quad K_{y,x}) = (2,1)$

Figure 4: Table of focal points

2.3 Third idea

The variables of a network are asynchronously updated toward their parameters, crossing at most one threshold.

The asynchronous updating is motivated by the fact that a threshold represents a very thin region of the real concentration space for each variable, thus, the probability that several variables cross their thresholds exactly at the same time is negligible. Consequently, when the network is in a state such that several variables can change (*i.e.*, such that several variables have a concentration level belonging to an interval which is different from the interval pointed by the current parameter $K_{...}$), there are as many possible next states as such variables. From such a state, the system can choose to modify any one of these variables.

For example, Figure 5 shows the state graph extracted from the network given in Figure 4. The left hand side of the figure shows what would happen if we followed a naive synchronous updating that would reflect the table of focal points: the situation would be biologically incredible for two reasons. The first reason, as already mentioned, is that x and y would be updated at the same time for example from the state (1,0). The second reason is that from the state (0,0), the variable x would cross two thresholds, which is contradictory with the fact that we are modelling a continuous change of concentration levels. The right hand side of the figure provides the correct abstract behaviours of the system.

It shows in particular that it is possible to reach the stable state (0,1) from the initial state (0,0).

2.4 The boolean framework

The boolean framework was the first framework introduced by René Thomas. It modifies the first idea as follows: *the product of a gene can be "present" or "not present" in the cell*. It means that there is only one threshold for each gene; the two other ideas (the parameters K and the asynchronous state graph) remain unmodified.

Let us consider the "type 1 incoherent feedforward loop" introduced in Figure 1. As we are in the boolean framework, every threshold is equal to 1. Moreover, if we want that b needs the presence of its activator a *and* the absence of its inhibitor c to be synthesized, then a unique choice is possible to make all the interactions of the graph functional, see the parameter valuation in Figure 6. Remember that, in the $K_{b...}$ parameters, the subscript c means that c does not pass the threshold, as it is an inhibitor of b. Lastly, the variable a being the entry point of the feedforward pattern, we do not consider K_a yet.

The question that we will address on this example all along the article is the following: *what shall be the behaviour of b in response to the input signal offered by a ?*

Obviously, if a is equal to 0 for a sufficiently long time, both b and c will also be equal to 0, because b and c need a as a resource in order to reach the state 1; see Figure 7. Let us assume that the signal a goes from 0 to 1. Then, the current state will move to $(a = 1, b = 0, c = 0)$: the square situated at the lower left corner of the plan $a = 1$ of Figure 7. The new stable state is $b = 0$ and $c = 1$ but, due to the asynchronous semantics, there are two different paths from the current state: either we go directly to the stable state and b remains constantly equal to 0, or we follow the other path where b is transitorily equal to 1, before being inhibited by c.

Under which conditions will b always signal the presence of a *via* a transitory production ? May-be the conjunction of resources for the variable b is not

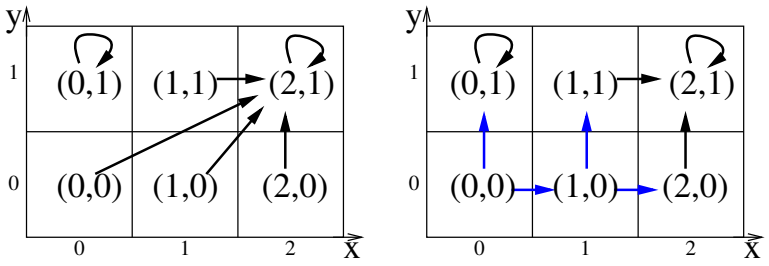
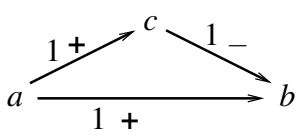


Figure 5: Synchronous and asynchronous state graphs



$$K_b = 0 \quad K_c = 0$$
$$K_{b,a} = 0 \quad K_{c,a} = 1$$
$$K_{b,c} = 0$$
$$K_{b,ac} = 1$$

Figure 6: Boolean feedforward

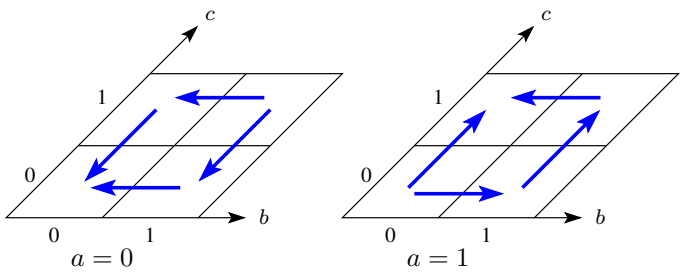


Figure 7: Asynchronous state graph for type 1 incoherent feedforward loop

optimal, e.g., would a disjunction be better ? or any other values for the $K_{...}$ parameters ?

This is more generally the usual question of identification of the parameter values, according to some biologically known behaviours or some hypothetical behaviours. Here, the example would be small enough to enumerate all the possible parameter values, to generate the state graphs and study for each of them the answer of b. It is of course not the case when addressing real size gene networks and so, formal methods from computer science are required to perform computer-aided identification of parameters.

2.5 Temporal logic and automatic model checking

Temporal logics are languages that allow us to formalize biologically known behaviours or hypothetical behaviours in such a way that computers can automatically check if a model exhibits those behaviours or not. The building blocks of a temporal logic are atoms, connectives and temporal modalities. Let us here consider the Computation Tree Logic [12, 17], CTL for short, which is the most common temporal logic:

- Atoms in our case are simple statements about the current state of a variable of the network. For example equalities (*e.g.*, $x = 2$) or inequalities (e.g., $x \leq 1$ or $y > 1$).
- Connectives are the standard connectives: negation (*e.g.*, $\neg(x = 0)$ is the negation of the atom $x = 0$), conjunction (*e.g.*, $(x = 0) \wedge (y > 1)$),

disjunction (*e.g.*, $(x = 0) \vee (y > 1)$), implication (*e.g.*, $(x = 0) \Rightarrow (y > 1)$), and so on.

- Temporal modalities are combinations of two types of information:
  - Quantifiers: a formula can be checked with respect to all possible choices of paths in the asynchronous state graph (universal quantifier, denoted by A), or one can check if it exists at least one path choice such that the formula is satisfied (existential quantifier, denoted by E).
  - Discrete time elapsing: a formula can be checked at the next state (letter X), in some future state which is not necessarily the next one (letter F), in all future states (letter G) and a formula can be checked until another formula becomes satisfied in the future (letter U).

In short:

first character	second character
A = for **A**ll path choices	X = ne**X**t state
	F = for some **F**uture state
E = there **E**xists a choice	G = for all future states (**G**lobally)
	U = **U**ntil

For example, the formula $((x = 0) \wedge (y > 0)) \Longrightarrow A[(x = 0)U(y = 0)]$ means that, starting from an initial state where $x = 0$ and y is strictly positive, there will be a state in the future such that $y = 0$ and meanwhile, x will remain equal to 0, whatever the choice of path. More generally, Figure 8 summarizes the CTL semantics with the following conventions: we start from an arbitrary initial state that constitutes the root of the tree; a blue arrow means that φ becomes true in the target of the arrow; a green arrow means that φ is not satisfied in the target of the arrow; a red arrow means that ψ is satisfied both in the source and in the target of the arrow.

One of the main advantages of CTL is that there are very efficient model checkers, see for example [9]. A model checker is an algorithm that takes as inputs a CTL formula and a state graph, and furnishes as output the subset of states that satisfy the formula.

Model checking can be used to identify the parameters that are compatible with the known or hypothetical behaviours [7]. SMBioNet is a software platform where we can enter the influence graph between genes and where we can enter CTL formulas that describe the known behaviours: it automatically computes all the sets of parameter values that are compatible with both the graph and the behavioural properties. Technically, SMBioNet generates all the possible state graphs and performs model checking. Then, a model is proposed if and only if all its states satisfy all the behavioural properties.

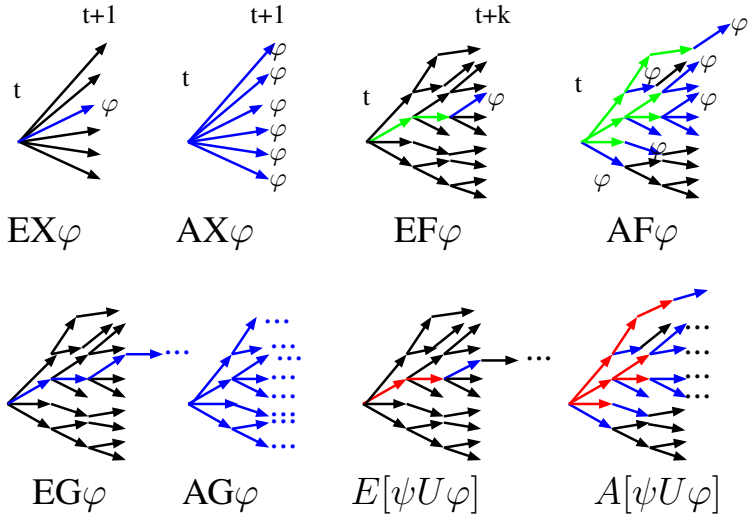


Figure 8: CTL modalities

For our feedforward example, the transitory activation of the gene b can be formalized in CTL as follows:

$$(b = 0 \wedge c = 0 \wedge AG(a = 1)) \Longrightarrow AF(b = 1 \wedge AXAG(b = 0))$$

It means that if the signal a becomes active when b and c are inactive, then b will become active in the future, and then it will become inactive.

We have also assumed that a is able to control c:

$$(a = 1 \wedge c = 0) \Longrightarrow EX(c = 1)$$
$$(a = 0 \wedge c = 1) \Longrightarrow EX(c = 0)$$

It means that when $a = 1$ (resp. $a = 0$) c can increase (resp. decrease). We use EX and not AX because the asynchrony can allow another variable to cross a threshold before c.

When submitting the formula to SMBioNet, we discover that there is no parameter values such that b always signals the switch of a by a transitory change of value: the direct path from $(a = 1, b = 0, c = 0)$ to $(a = 1, b = 0, c = 1)$ seems unavoidable whatever the values of the parameters.

Nonetheless, if we assume for instance that the *delay* for a gene to act on another gene is identical for all interactions in Figure 6, then a would start both the expression of b and c almost at the same time, and only after another delay, c will switch b to 0. So, b will always signal the switch of a.

The paradox comes from the fact that the standard Thomas' approach does not take delays into account. The parameters $K_{...}$ control only the functionality of combined interactions, not the delays. So, the asynchrony of variable

updates always gives the possibility for a to activate c and then for c to inhibit b before the direct activation of b by a will take place. More generally, it may also depend on the real initial state inside the square $(a = 1, b = 0, c = 0)$ and it may also depend on the relative production speeds of b and c. Indeed the standard Thomas' framework has a rough notion of *time*: it is reduced to the random scheduling of variable changes. This motivates the introduction of delays into the modelling framework.

2.6 Logic programming with constraints

Before discussing the different ways to introduce delays into the Thomas' framework, let us mention the importance of constraint solving for the parameter identification problem. The current platform SMBioNet exhaustively generates the possible state graphs and checks on them the temporal properties. When time delays will be introduced, they will of course constitute additional parameter values which will need to be identified as well. Delays shall be real numbers because time passes continuously, and consequently, an exhaustive enumeration of all the possible behaviours will become impossible. Temporal properties will induce constraints on both the Thomas' parameters and the time delay parameters, in such a way that the set of solutions will involve intervals of real time delays containing an infinity of points.

In the standard discrete framework of R. Thomas, L. Trilling has already proposed to use logic programming with constraints in order to identify the $K_{...}$ values [13]. More precisely, the method extracts all the parameter values that make possible a given set of observed paths in the state graph. The method has also been extended and implemented by F. Corblin [10] in the same research team, and the results are impressive. Provided that the temporal properties under consideration can be expressed *via* a finite number of paths of fixed length, a few seconds of computing time are needed for problems where SMBioNet needs several hours.

The idea is to specify, in the PROLOG language, the Thomas' asynchronous construction of the state graph, according to symbolic representations of the $K_{...}$ parameters. Then, by specifying that a given path exists in the state graph, PROLOG will generate the constraints on the parameters that permit each transition of the path. Lastly, constraint solving algorithms try to exhibit parameter values or to prove inconsistencies.

As an example, let us consider the path $(b = 0, c = 0) \rightarrow (b = 1, c = 0) \rightarrow (b = 1, c = 1) \rightarrow (b = 0, c = 1)$ in the plan $a = 1$ as in Figure 7. The first transition of the path generates the constraint $(K_{b,ac} > 0)$ because the variable b goes from 0 to 1 when a and (the absence of) c are resources of b. The two other transitions generate similarly $(K_{c,a} > 0)$ and $(K_{b,a} < 1)$ respectively. In order to ensure that b will always signal the presence of a *via*

a transitory switching on, one has to negate the existence of the path $(b = 0, c = 0) \rightarrow (b = 0, c = 1)$, which generates the negation of $(K_{c,a} > 0)$. Lastly, constraint solving will trivially prove that the resulting global set of constraints is inconsistent. The same computations must be done for all the paths that exhibit a transitory production of b.

Of course, both constraint programming methods and model checking methods give the same result and they both raise the same "delay paradox" mentioned previously. The simple chronologic notion of random scheduling of variable changes is not sufficient; we need an explicit notion of chronometric delays in the modelling framework.

3 Piecewise Linear Differential Equations

Since chronometric information is of great importance in the dynamics of the modelled system, it seems natural to come back to the framework of differential equations because differential systems make the time explicit. Moreover in this modelling framework, the trajectories are deterministic: from an initial state, the whole trajectory can be computed.

Nevertheless, parameters of the differential equations are generally not known and have to be determined. If we want to use knowledge on the time that takes a particular trajectory between two points, in order to determine unknown parameters, one has to explicite the relationship between elapsed time along a trajectory and parameters. Thus the differential equation system has to be solved. Generally, if the differential system has no particular shape, the symbolic solving of the differential system is not possible and the large number of variables makes appear additional difficulties. The computer tools which are useful for simulations of such systems are nevertheless not well adapted for this difficult task.

3.1 Piecewise Linear Differential Equations

To simplify this task, we can restrict the form of the differential system.
Snoussi [24] proposed to construct a piecewise linear differential equation system: with each qualitative situation (that is when interactions does not change) is associated a differential system which is easy to solve symbolically. The way to construct such a system of differential equations can be sketched as follows:

- With each node of the interaction graph is associated a variable of the differential equation. This variable represents the concentration of the associated protein.

- Each variable has a particular degradation rate. The degradation is supposed to be proportional to the concentration of the protein (greater the

concentration, greatest the degradation).

- Each variable has a synthesis rate which depends on the activity of its regulators (greater the number of activators, greatest the synthesis rate).
- Each predecessor of a variable (in the interaction graph) has an influence on the synthesis rate of the considered variable: if it is an activator, the synthesis rate is increased when the regulator has a concentration greater than the threshold associated with the interaction; if it is an inhibition, the synthesis rate is increased when the regulator has a concentration smaller than the threshold, see Figure 9.

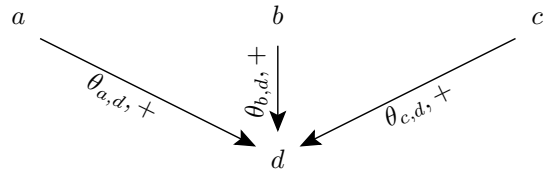


Figure 9: Example of a gene regulated by two activators and by one inhibitor.

The previous outline leads to the following differential equation system:

$$\frac{d}{dt}x_i = \left(k_0^i + \sum_{j \in A(i)} k_j^i \times \mathbb{1}_{[x_j > \theta_{j,i}]} + \sum_{j \in I(i)} k_j^i \times \mathbb{1}_{[x_j < \theta_{j,i}]} \right) - \gamma_i x_i$$

where $A(i)$ (resp. $I(i)$) is the set of activators (resp. inhibitors) of i, and $\mathbb{1}_{[\text{condition}]}$ is equal to 1 if the condition is satisfied and equal to 0 otherwise[1]. The first term is the synthesis rate which can be decomposed into three part:

- k_0^i which is the basal synthesis rate,
- the contribution of activators (each activator contributes to the synthesis rate when its concentration is greater than some threshold) and
- the contribution of inhibitors (each inhibitor contributes to the synthesis rate when its concentration is less than some threshold).

[1]Let us remark that when a concentration is on a threshold, the contribution of the associated action is not taken into consideration. It would be better to consider that the differential equation is not defined on thresholds: one does not know whether the regulation takes place or not. If one is interested in the precise behaviour of the system on thresholds, one has to embed such a differential equation into the framework of differential inclusions [14]. This work has already been done in the context of gene regulatory networks [11].

Finaly $\gamma_i x_i$ represents the degradation. Such a differential system is called a Piecewise Linear Differential Equation system, PLDE for short.

The previous differential equation system is based on the qualitative contribution of each regulator. Unfortunately, even if the additivity of contributions is not put into question, the contribution of a regulator is not a discontinuous step function. To improve the model, the set fonction $\mathbb{1}_{[x>\theta]}$ (resp. $\mathbb{1}_{[x<\theta]}$) can be replaced by a Hill function:

$$H_+(x) = \frac{x^n}{\theta^n + x^n} \qquad \left(\text{resp. } H_-(x) = \frac{\theta^n}{\theta^n + x^n}\right)$$

where n is the parameter of the Hill function which controls its roughness.

3.2 Coherence between PLDE and discrete models

Such a differential equation system has a deep relationship with the discrete models of Section 2. Let us first remark that the thresholds allow a discretization of the phase space: ranking the thresholds $\{\theta_{j,i} | i \text{ is a possible target of } j\}$ for each variable j allows one to split the concentration space of j into different subdomains numbered from 0 to b_j. The discretization of the continuous phase space is then defined by associating with each concentration state s, the discrete vector characterizing the subdomain of s. Thus the parameter of the discrete model $K_{i,\omega}$ is the discretization of the coordinate i of the equilibrium point of the differential system associated with the situation where ω is the set of regulators contributing to the synthesis rate:

$$\begin{array}{c}\text{i}^{th} \text{ coordinate of the} \\ \text{equilibrium point}\end{array} = \left(\frac{k_0^i + \sum_{j \in A(j) \cap \omega} k_j^i + \sum_{j \in I(j) \cap \omega} k_j^i}{\gamma_i}\right)$$

If each contribution to the synthesis rate $k_{i,\omega}$ is positive and if, for each i and each ω, $K_{i,\omega}$ is equal to the discretization of the ith coordinate of the equilibrium point, then there exists a transition from discrete state s_1 to s_2 if and only if there exists a trajectory of the differential system starting from the domain associated with s_1 and going to the threshold separating this domain from the domain associated with s_2 [24].

3.3 A feedforward loop controled by a positive auto-regulation

To study the behavior of an incoherent type 1 feedforward loop, we first consider that the action of the transcription factor a does not change, that is, that its level of concentration does not cross the threshold of one of its interactions. The simplest way to study such a system is to consider that the transcription factor a is also a regulator of itself, see Figure 10. This positive auto-regulation

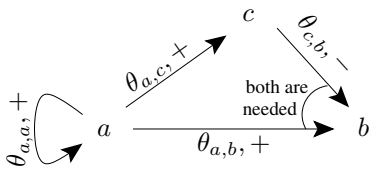


Figure 10: Incoherent type 1 feedforward loop combined with a positive autoregulation of a.

leads to multi-stationarity [28] of a: if a is present (resp. absent), it remains present (resp. absent). As in Section 2.4 we suppose that b needs the presence of its activator a and the absence of its inhibitor c to be synthesized. Mathematical modelling suggested that the I1-FFL can show two dynamical features [20] amongst whose a transient pulse of expression of b. To verify this possible behaviour, let us build the corresponding differential equation system:

$$\left\{\begin{array}{lcl} \frac{da}{dt}(t) & = & k_0^a + k_a^a \mathbb{1}_{[a>\theta_{a,a}]} - \gamma_a.a(t) \\ \frac{db}{dt}(t) & = & k_0^b + k_a^b \mathbb{1}_{[a>\theta_{a,b}]} + k_c^b \mathbb{1}_{[c<\theta_{c,b}]} - \gamma_b.b(t) \\ \frac{dc}{dt}(t) & = & k_0^c + k_a^c \mathbb{1}_{[a>\theta_{a,c}]} - \gamma_c.c(t) \end{array}\right.$$

Nevertheless this differentential system does not express that b needs the presence of its activator a *and* the absence of its inhibitor c to be synthesized. It has to be modified to take into account the condition under which the activation of b is effective: $[(a > \theta_{a,b}) \wedge (c < \theta_{c,b})]$:

$$\left\{\begin{array}{lcl} \frac{da}{dt}(t) & = & k_0^a + k_a^a \mathbb{1}_{[a>\theta_{a,a}]} - \gamma_a.a(t) \\ \frac{db}{dt}(t) & = & k_0^b + k_{ac}^b \mathbb{1}_{[(a>\theta_{a,b})\wedge(c<\theta_{c,b})]} - \gamma_b.b(t) \\ \frac{dc}{dt}(t) & = & k_0^c + k_a^c \mathbb{1}_{[a>\theta_{a,c}]} - \gamma_c.c(t) \end{array}\right.$$

where k_{ac}^b is the contribution to the synthesis rate of b due to the presence of a and the absence of c.

The global behaviour of such a system is driven by the values of parameters. For example, for some values of parameters, the protein b is synthesized before the action of the inhibitor takes place, see Figure 11-left, whereas for some other values of parameters, the synthesis of b is not so visible, see Figure 11-right. Let us remark that in both cases, the equilibrium point of variable b when a is present and c absent is then same: $\frac{k_0^b + k_{ac}^b}{\gamma_b} = \frac{40}{2} = \frac{5}{0.25} = 20$. Both models lead to the same discrete model.

Thus the identification of parameters becomes a crucial step also in the PLDE modelling framework, because a variation of parameters can lead to different qualitative behaviours, see Figure 11. Moreover the values of kinetic

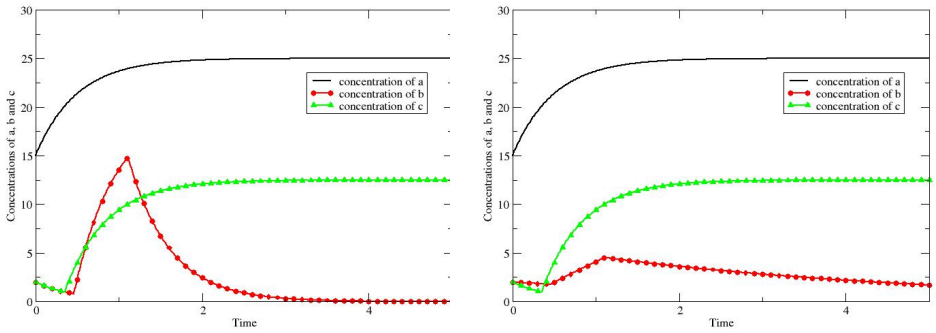


Figure 11: Feedforward loop controled by a positive auto-regulation: according to kinetic parameters, b can be activated before the effect of the inhibition of c or not. Left: $\gamma_b = 2$ and $k^b_{ac} = 40$ Right: $\gamma_b = 0.25$ and $k^b_{ac} = 5$. Other parameters are identical for both simulations: $\theta_{a,a} = 10.0$, $\theta_{a,b} = 21.0$, $\theta_{a,c} = 20.0$, $\theta_{c,b} = 10.0$, $\gamma_a = \gamma_c = 2$, $k^a_0 = k^b_0 = k^c_0 = 0.0$, $k^a_a = 50.0$, $k^c_a = 25$. Initial state is $(15, 2, 2)$.

parameters $k^{\cdot}_{\dots}$ are mandatory to deduce the sequence of domains the trajectory passes through, which are also necessary to compute the time that takes a trajectory passing through such a sequence of domains.

The first idea to overpass the *parameter identification problem* is to grope for parameters until a set of parameters leads to a behaviour compatible with available information about the trajectories. After having found a valuation of parameters, simulations of the mathematical model are performed (several of them under perturbations) in order to evaluate its robustness, that is its ability to maintain its functions against internal and external perturbations [18]. Indeed, since robustness is one of the fundamental characteristics of biological systems and has been demonstrated many times experimentally [19], the evaluation of the robustness of the PLDE model is a indicator of its validity. Nevertheless the evaluation of the robustness of the PLDE model does not validate completely the model.

3.4 A feedforward loop controled by a negative loop

We now study the behavior of the incoherent type 1 feedforward loop when the transcription factor a oscillates. The simplest way to study such a system is to consider the interaction graph made of the incoherent type 1 feedforward loop and of the negative loop $(a \rightleftarrows a')$ containing the transcription factor a, see Figure 12. The negative feedback loop leads to oscillations of a and a' under

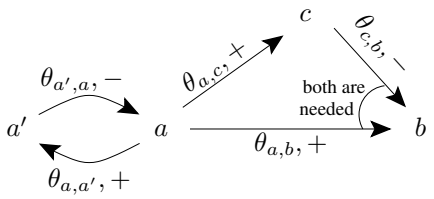


Figure 12: Incoherent type 1 feedforward loop combined with a negative loop.

some conditions, in such a case, we say that the circuit is *functional*. When the period of oscillation of a and a' is sufficiently small, neither b nor c is able to switch-on during a unique period. But if degradation rate is also sufficiently weak, several period can lead to the activation of b, c or both b and c. The PLDE model can be easily written:

$$\left\{\begin{array}{lcl} \frac{da}{dt}(t) & = & k_0^a + k_{a'}^a \mathbb{1}_{[a<\theta_{a',a}]} - \gamma_a.a(t) \\ \frac{da'}{dt}(t) & = & k_0^{a'} + k_a^{a'} \mathbb{1}_{[a>\theta_{a,a'}]} - \gamma_{a'}.a'(t) \\ \frac{db}{dt}(t) & = & k_0^b + k_{ac}^b \mathbb{1}_{[(a>\theta_{a,b})\wedge(c<\theta_{c,b})]} - \gamma_b.b(t) \\ \frac{dc}{dt}(t) & = & k_0^c + k_a^c \mathbb{1}_{[a>\theta_{a,c}]} - \gamma_c.c(t) \end{array}\right. \quad (1)$$

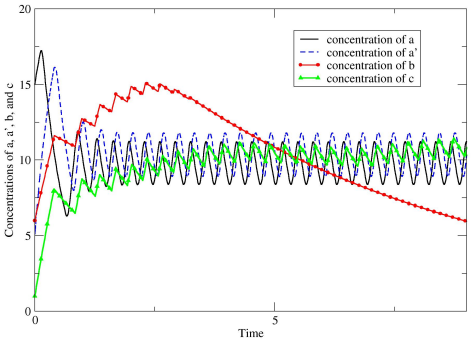
This differential system can lead to subtile behabiors. Let us first suppose that oscillations of a and a' are much faster than the increasing of b and c. Thus the order of activation of genes b and c, is intermittent and several behaviours can be obtained according to relative values of the synthesis and degradation rates:

1	neither accumulation of b nor accumulation of c
2	accumulation of b but no accumulation of c
3	no accumulation of b but accumulation of c
4	accumulation of b and c, but c is activated before b
5	accumulation of b and c, but b is activated before c

Table 1: The different possible behaviours of the feedforward loop controled by a negative loop.

Figure 13 shows the evolution of concentrations of 4 variables for some parameter values. It is clear that this choice of parameter values corresponds to the situation where both variables b and c increase because of the intermittent order due to a: synthesis rate of b (resp. c) when a does activate b (resp. c) is sufficiently high to allow, after a total oscillation period of a, a little accumulation even after the second phase of the cycle when the activation order is off. In other word, during each oscillation cycle of a, the system creates more b than it degrades b. Such an accumulation of b and c are due to

low degradation rate but such values are nevertheless credible: the permease in the lactose operon system is known to be degradated very slowly. Moreover,



Thresholds	Degradations
$\theta_{a,a'} = \theta_{a',a} = 10$	$\gamma_a = 2$
$\theta_{ac} = 10.6$	$\gamma_{a'} = 2$
$\theta_{a,b} = 10.5$	$\gamma_b = 0.15$
$\theta_{c,b} = 9.5$	$\gamma_c = 0.5$
Synthesis rates	Initial state
$k_0^a = k_0^{a'} = k_0^b = k_0^c = 0,$	$a_0 = 15$
$k_{a'}^a = k_a^{a'} = 50$	$a'_0 = 6$
$k_{ac}^b = 15.0$	$b_0 = 2$
$k_a^c = 20$	$c_0 = 2$

Figure 13: Incoherent type 1 feedforward loop combined with a negative loop: the negative loop generates oscillations which allow b and c to accumulate.

in Figure 13, b increases faster than c. Thus b becomes present but when c becomes greater than the threshold of its action on b, variable b begins to decrease and will no more be activated. Such a behaviour corresponds to the situation 5 of the previous table.

Unfortunately, as said in the introduction, there does not exist an automated method to extract properties of kinetic parameters which have to be fullfiled to allow the system to present any known dynamical property. We then set out in the next section a hybrid framework based on the discrete one which tries to mimics the different behaviours of PLDE systems.

4 First hybrid modelling approach due to R. Thomas

To try to automate the parameter identification step for a timed model of a gene regulatory network, it seems natural to propose to build a timed version of the discrete approach since this discrete framework can be viewed as a discretization of the PLDE framework. The refined modelling is based on the use of delays of activation / inhibition to specify which variable is faster affected by a change of its regulators. To be more precise, when an order of activation / inhibition rises, the biological machinery starts to increase or to decrease the corresponding protein concentration, but this action takes time. Thus the differences between the values of delays of activation / inhibition lead to decrease the non-determinism.

4.1 Qualitative states and clocks

This idea dates back to the book of Thomas and d'Ari [27]: with each variable is associated a clock which measures the elapsed time, and each transition needs some delay to be passed over. The simulation of such a model can then be sketched as follows:

1. The initial state is made of a discrete state and a initialisation of clocks (generally each clock is set to 0).
2. According to the current discrete state, the clocks associated with variables whose focal point allows them to evolve (that is whose focal point is placed outside the domain), run simultaneously at the same speed.
3. The next fired discrete transition is given by the clock which first reaches its associated delay. If two delays are equal, that is if two clocks reach their delays at the same time, non-determinism remains and several discrete transitions can be fired. In such a case, choose at random a possible transition.
4. In the new state, some clocks are reset: the clock which allowed the transition is reset to zero, but also each clock for which the order has changed. For example, if in the previous state, the variable a was subject to an decreasing order, but in the new state, it is subject to an increasing order, its associated clock is reset also to 0.
5. Repeat steps 2, 3 and 4.

4.2 A feedforward loop controled by a positive auto-regulation

Depending of the delays associated with transitions, two behaviours can be simulated: the first one allows the switch-on of variable b, while the second does not allow it. Let us consider the boolean network described in Section 2.4 completed by the auto-regulation of a. The functionality of the auto-regulation of a, which does not allow a to evolve from its initial state, leads to following values of parameters concerning variable a:

$$K_a = 0 \qquad K_{a,a} = 1$$

Other parameters are the same, see Figure 6. Because a is not able to evolve from its initial state, the state graph is the one of Figure 7. The initial state is the boolean state $(a = 1, b = 0, c = 0)$ combined with an initialization of clocks where each clock is set to zero. On one hand, if the delay mandatory to activate c is less than the delay mandatory to activate b, then b will never be switched on because the inhivitor c becomes rapidly effective. On the other

hand, if the delay to activate c is greater, c gives b the time to be switched on before becoming an effective inhibitor.

We proposed in [2] a formalisation of such a modelling approach which is based on two types of parameters, $d_v^+(x)$ and $d_v^-(x)$, which represent the time delays required to change the expression level of a variable v from level x to $x+1$ and from level x to $x-1$, respectively, as shown in Figure 14. Then,

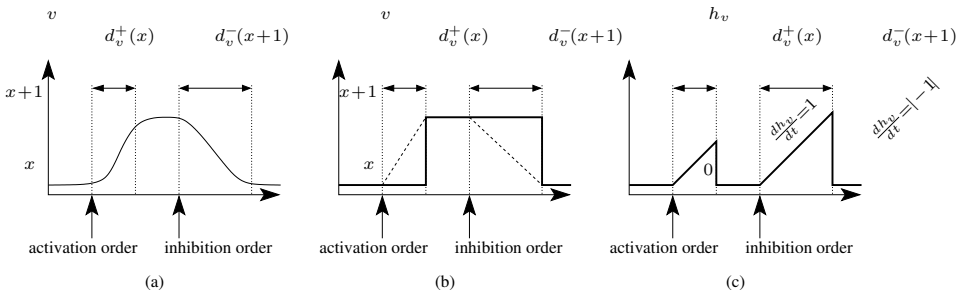


Figure 14: Evolution of a gene's expression (a), its schema in the discrete model (b) and in its extension with time delays (c).

we add to each variable v a continuous clock h_v whose speed at state μ is 1 (when variable v can evolve) or 0 (if it cannot). At a given qualitative state μ, if the concentration of v is increasing (resp. decreasing), then, when h_v reaches $d_v^+(\mu(v))$ (resp. $d_v^-(\mu(v))$), the level of v becomes $\mu(v)+1$ (resp. $\mu(v)-1$) and the clock h_v is reset.

The temporal model described above belongs to the class of the so-called *stopwatch automata* [8] which is a specific type of linear hybrid automata [4, 5], LHA for short. LHA are finite state automata augmented with real variables whose values evolve continuously in a discrete state. Whereas the values of the continuous variables can be affected by discrete transitions between discrete states, evolutions of continuous variables are lines inside a discrete state. Linear hybrid automata can be subject to a reachability analysis. However, in general, the reachability problem for linear hybrid automata is undecidable [25].

In such a modelling framework, the parameter identification problem still remains the cornerstone of the approach. The determination of discrete parameters (the $K_{v,\omega}$) can be driven by model checking as shown in section 2.5. It then remains to identify the delays. Since time delays are real numbers, it cannot exist any enumeration method (SMBioNet-like) which tries all possible combinations of delay values and retains only those which are coherent with knowledge about the behaviour. One then have to turn to constraints in order to express the conditions under which the known properties are satisfied by the model.

Let us focus on the example of the feedforward loop controled by a positive auto-regulation.

1. Let us suppose that available knowledge about the system allows one to state that when a is on, b is switched on before c. Afterwards b is switched off (because c becomes present) n minutes after the switch-on of a. In other words, the discrete path $(1,0,0) \rightarrow (1,1,0) \rightarrow (1,1,1) \rightarrow (1,0,1)$ has to be possible in the model with delays.

   - $(1,0,0) \rightarrow (1,1,0)$ leads to the constraint

     $$\delta_b^+(0) < \delta_c^+(0)$$

   - Moreover the time that takes a trajectory from discrete state $(1,0,0)$ to $(1,0,1)$, is $\delta_b^+(0) + (\delta_c^+(0) - \delta_b^+(0)) + (\delta_b^-(1))$. Then we also have the following constraint:

     $$\delta_c^+(0) + \delta_b^-(0) = n \text{ minutes.}$$

2. Let us now consider that the switch-on of b does not occur. The deduced constraint becomes:

   $$\delta_c^+(0) < \delta_b^+(0)$$

Such kind of constraints can be automated by the use of some computer science tools dedicated to analysis of linear hybrid automata. For example we used HyTech [16] for two purposes: (1) to find automatically all paths from a specified initial state to another one and (2) to synthesize constraints on the delay parameters in order to follow any specific path.

This modelling framework then seems to allow the modeller to take into account information about observed time. Indeed the parameter identification can be decomposed into two parts: the valuation of discrete parameters can be found using an exhaustive approach like SMBioNet, and, the delay parameters can be found using HyTech which allows the building of constraints.

Nevertheless, such a modelling framework present a little drawback: the succession of intermittent orders of synthesis of a variable cannot lead to its global increase. This drawback is explicit in the following example.

4.3 A feedforward loop controled by a negative loop

Let us recall that a and a' oscillate with a period which is much less than the delays mandatory for the swich-on of variables b and c. Concentrations of b and c are then increasing but at each cycle of a, the counter-order of decreasing of b (resp. c) resets the clock h_b (resp. h_c) before the clock has

reached the threshold $\delta_b^+(0)$ (resp. $\delta_c^+(0)$). Then neither h_b nor h_c will reach the threshold leading to the switch-on of the corresponding variable. This modelling framework can only represent the situation where neither b nor c can switch-on, case 1 of Table 1.

Thus this modelling framework makes possible the automation of parameter identification, and allows the distinction of two different behaviours (b can be switched-on or not) amongst the 5 possible ones. Nevertheless, it does not allow the representation of accumulation.

5 Product of automata: an alternative approach

The HyTech model checker performs *symbolic model checking* on automata and we have shown that this extension of model checking allows for the extraction of parameter constraints from some given paths. Another way to use symbolic model checking in order to identify the parameters and the delays of a gene regulatory network is to perform *products of automata*. The advantage of this kind of approach is that the computation of a product of automata does not only furnish a resulting automaton; it also systematically labels the states and the transitions of the automaton by some formulas that define the conditions under which the transitions can be fired. Then, provided that we adopt an adequate "hybrid" temporal logic, there are model checking algorithms able to manage symbolic values for some parameters. They compute the constraints under which a given temporal formula is satisfied.

Using the UPPAAL model checker

In [23], Heike Siebert and Alexander Bockmayr made use of products of automata in order to formalize a hybrid modelling framewok for delays inspired by the approach of René Thomas. The automata that play the role of state graphs in this framewok are rather heavy, but they should be considered as purely technical mathematical objects, which will be submitted to UPPAAL [6].

The main idea is the following. For each variable v of the network, there is a clock called h_v, and for each possible discrete state of this variable, there are three possible behaviours with respect to delays:

- either the parameter $K_{v,\omega}$ (where ω is the set of resources of v according to the current state of the system) is greater than the current value of v, in which case the clock h_v measures the time of increasing of v;
- either the parameter $K_{v,\omega}$ is lower than the current value of v, in which case the clock h_v measures the time of decreasing of v;

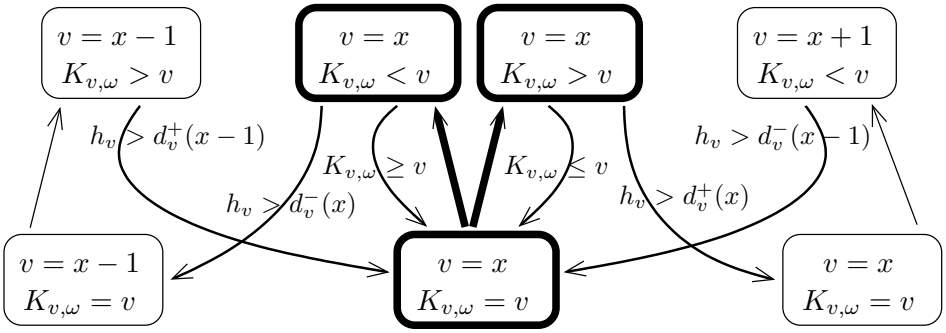


Figure 15: Timed automaton for one level

- or the parameter $K_{v,\omega}$ is equal to the current value of v, in which case the clock h_v is off.

This principle is reflected by an "atomic" automaton structure containing three discrete states (called *locations* in the timed automata framework), as shown in the bold part of Figure 15. The central location is intuitively the default location for the state $v = x$ and every transition that changes the value of the variable v goes to this central location. Then, if $K_{v,\omega} = v$ is false, the suitable bold transition goes *immediately* to the consistent location (either $K_{v,\omega} < v$ or $K_{v,\omega} > v$) and the clock h_v starts from 0.

The product of these atomic automata is managed in such a way that the set of resources ω is properly computed in the product automaton. The formulas are rather complex, but they simply reflect the formal definition of the Thomas' framework. Lastly, as shown in the figure, if the clock h_v reaches its limit delay $d_v^+(x)$ (resp. $d_v^-(x)$) then the transition to $v = x + 1$ (resp. $v = x - 1$) is fired, and similarly, if some other variable change induces a different comparison of $K_{v,\omega}$ with respect to x, then the location is pulled back to the central one.

This technical stuff being done, UPPAAL can be used in order to extract the constraints generated by some temporal formulas, which can for example reflect knowledge on the biological system about time delays to go from one state to another state. Let us remark that this framework still does not treat accumulation, because when the location is pulled back to the central one, it resets automatically the clock to 0.

6 Hybrid models inspired by PLDE

We saw that situations where accumulation plays a crucial role for the global behaviour are difficult to take into consideration. Nevertheless we want to

overpass these difficulties and propose a new hybrid modelling framework which takes into account accumulation.

The first attempt to propose such a hybrid modelling framework [1] was based on the discrete model. With each discrete state, is associated a temporal zone, which makes hybrid the models. The temporal zone is defined as a hypercube whose dimension is the number of variables. The length of the hypercube associated with state μ in the direction v is $d_v^-(\mu) + d_v^+(\mu)$: it corresponds to the sum of the time mandatory to pass to the level $\mu_v + 1$ under increasing order and the time mandatory to pass to the level $\mu_v - 1$ under decreasing order. Because of the presence of two delays associated with a same domain and a same variable, accumulation can be represented but the decreasing and the decreasing of a variable subject to accumulation take place at the same speed, this drawback has been discussed in [1], see Figure 13 inside.

6.1 From PLDE to hybrid models

Since PLDE modelling framework is able to represent such accumulations, we present in this section yet another hybrid modelling framework based on PLDE which partially allows the building of constraints leading to a particular behaviour. The only new fundamental idea is to express a relationship between delays of the hybrid model and the PLDE model: the delay $d_v^+(\mu)$ (resp. $d_v^-(\mu)$) is an approximation of the time necessary to variable v to cross the domain from the lower bound to the upper bound (resp. from the upper bound to the lower one).

In other words, if the PLDE is known, it becomes easy to build the hybrid model since

- the thresholds defining the discretization of the PLDE are given,
- the parameters $K_{...}$ are the discretization of equilibrium points,
- the delays parameters are deducible from the PLDE.

More interesting is the inverse translation: If a hybrid model is supposed to represent a system, is it possible to construct a PLDE system whose behaviours are coherent with the possible paths in the hybrid model ? To answer this question, the work of Snoussi [24] has to be done again in the hybrid context: Snoussi has shown that it is possible to build from a discrete model, a PLDE system whose discretization is the discrete model.

Intuition. Let us consider the differential system modelling the feedforward loop controled by a positive auto-regulation:

$$\left\{\begin{array}{lcll} \frac{da}{dt}(t) & = & 50 \times \mathbb{1}_{[a>10]} & -2 \times a(t) \\ \frac{db}{dt}(t) & = & 40 \times \mathbb{1}_{[(a>21)\wedge(c<10)]} & -2 \times b(t) \\ \frac{dc}{dt}(t) & = & 25 \times \mathbb{1}_{[a>20]} & -2 \times c(t) \end{array}\right.$$

A simulation of this differential system is shown in Figure 11-left. In order to consider two different qualitative level of b, we introduce a threshold $\theta_b = 10$: if concentration of b is less than θ_b, b is said absent, otherwise it is said present. Let us suppose that a is greater than the threshold $\theta_{a,a}$. 16 domains have to be considered since a can take 4 qualitative values (less than $\theta_{a,a} = 10.0$, between $\theta_{a,a}$ and $\theta_{a,c} = 20.0$, between $\theta_{a,c}$ and $\theta_{a,b} = 21.0$, or greater than $\theta_{a,b}$), b and c can take 2 qualitative values (greater or less than θ_b and $\theta_{c,b}$).

1. In the domain $(1, 0, 0)$, a is increasing towards $50/2 = 25$

$$\left\{\begin{array}{lcll} \frac{da}{dt}(t) & = & 50 & -2 \times a(t) \\ \frac{db}{dt}(t) & = & & -2 \times b(t) \\ \frac{dc}{dt}(t) & = & & -2 \times c(t) \end{array}\right.$$

   The delay $d_a^+((1, 0, 0))$ is deduced from the solution of the differential equation: $a(t) = \frac{50}{2} - (\frac{50}{2} - a(0))e^{-\gamma_a t}$. The delay is the time necessary for the solution to go from the left boundary of the domain to the right one. Then $a(0) = \theta_{a,a} = 10$, and $a(t) = \theta_{a,c} = 20$.

$$d_a^+((1, 0, 0)) \quad = \quad -\frac{1}{\gamma_a} \ln\left(\frac{\frac{50}{2} - \theta_{a,c}}{\frac{50}{2} - \theta_{a,a}}\right) = -\frac{1}{2}\ln(\frac{1}{3}) = 0.55 \quad (2)$$

   Both other variables stays in the domain, then delays are not significant.

2. In the domain $(2, 0, 0)$, a and c are subject to increasing order towards $50/2 = 25$ and $25/2 = 12.5$ respectively.

$$\left\{\begin{array}{lcll} \frac{da}{dt}(t) & = & 50 & -2 \times a(t) \\ \frac{db}{dt}(t) & = & & -2 \times b(t) \\ \frac{dc}{dt}(t) & = & 25 & -2 \times c(t) \end{array}\right.$$

   The delays $d_a^+((2, 0, 0))$ and $d_c^+((2, 0, 0))$ are deduced from the solution of the differential equation: $a(t) = \frac{50}{2} - (\frac{50}{2} - a(0))e^{-\gamma_a t}$ and $c(t) = \frac{25}{2} - (\frac{25}{2} - c(0))e^{-\gamma_c t}$. The delay is the time necessary for the solution to go from the left boundary of the domain to the right one.

- for variable a: $a(0) = \theta_{a,c} = 10$, and $a(t) = \theta_{a,b} = 21$.

$$d_a^+((2,0,0)) = -\frac{1}{\gamma_a} \ln \left(\frac{\frac{50}{2} - \theta_{a,b}}{\frac{50}{2} - \theta_{a,c}} \right) = 0.11 \quad (3)$$

- for variable c: $c(0) = 0$, and $c(t) = \theta_{c,b} = 10$.

$$d_c^+((2,0,0)) = -\frac{1}{\gamma_c} \ln \left(\frac{\frac{25}{2} - \theta_{c,b}}{\frac{25}{2} - 0} \right) = 0.84 \quad (4)$$

Then in the hybrid model, unless if c has accumulated before, a will pass the threshold $\theta_{a,b}$ before c will cross $\theta_{c,b}$.

3. In the domain $(3,0,0)$, b and c are subject to increasing order towards $40/2 = 20$ and $25/2 = 12.5$ respectively.

$$\begin{cases} \frac{da}{dt}(t) = 50 \quad -2 \times a(t) \\ \frac{db}{dt}(t) = 40 \quad -2 \times b(t) \\ \frac{dc}{dt}(t) = 25 \quad -2 \times c(t) \end{cases}$$

The solutions of the differential equation system are:
$b(t) = \frac{40}{2} - (\frac{40}{2} - b(0))e^{-\gamma_b t}$ and $c(t) = \frac{25}{2} - (\frac{25}{2} - c(0))e^{-\gamma_c t}$.
The delays are the times necessary for the solution to go from the left boundary of the domain to the right one.

- for variable b: $b(0) = 0$, and $b(t) = \theta_b = 10$.

$$d_b^+((3,0,0)) = -\frac{1}{\gamma_b} \ln \left(\frac{\frac{40}{2} - \theta_b}{\frac{40}{2} - 0} \right) = 0.34 \quad (5)$$

- for variable c: $c(0) = 0$, and $c(t) = \theta_{c,b} = 10$.

$$d_c^+((3,0,0)) = -\frac{1}{\gamma_c} \ln \left(\frac{\frac{25}{2} - \theta_{c,b}}{\frac{25}{2} - 0} \right) = 0.84 \quad (6)$$

Other delays are deduced similarly.

6.2 Sketch of the hybrid model

To go further, one needs to describe the evolutions of the hybrid model. In fact, with each domain is associated a temporal zone which is also defined as a hypercube whose dimension is the number of variables. According to the position of the focal point, we split the temporal zone into several subzones. Let us consider a discrete state $\mu = (\mu_i)_{i \in V}$:

- if $K_{i,\omega_i(\eta)} = \mu_i$, the coordinate i of the temporal zone is divided into 3 parts: the part where the concentration of i has to increase in order to reach the coordinate i of the focal point (the clock associated with i continues to increase until a delay denoted $d_i^+(\eta)$), the part where the concentration of i has to decrease in order to reach the coordinate i of the focal point (the clock continues to decrease until a delay denoted $d_i^-(\eta)$), and the part where the concentration of i has reached the coordinate i of the focal point (the clock is stopped).

- if $K_{i,\omega_i(\eta)} > \mu_i$ (resp. $< \mu_i$), the coordinate i of the temporal zone is not divided, since in all cases, the concentration of i has to increase (resp. decrease). The delay $d_i^-(\eta)$ (resp. $d_i^+(\eta)$) is set to 0.

The states of the hybrid model are couples $(\eta, (c_i)_{i\in V})$ where η is a qualitative state and $c_i \leq d_i^-(\eta) + d_i^+(\eta)$. Evolutions inside a qualitative state are easy to describe: the system evolves linearly until a boundary is reached: if the reached boundary corresponds to a subzone where variable i does not evolve anymore, then the clock associated with variable i stopped.

The description of the transition between two temporal zones is a little more tricky. If the reached boundary is a external face of the temporal zone, there is a qualitative jump from the current discrete state to the next state. The clock of the variable which has changed is reset in order to be coherent with the new qualitative state, and other clocks are modified to preserve the proportion of the concentration space which has already been crossed. Some particular situations lead to tricky rules explaining, for example, what is the trajectory when one successor of state μ_1 is μ_2 and one successor of state μ_2 is μ_1 (see the notion of black wall in [11]). The precise definition of this hybrid model can be found in [15].

Let us notice that delays associated with i seem to depend on the current qualitative state. Nevertheless for all qualitative states where the regulators of i are identical, the differential equation for variable i is the same. Thus, delays associated with i depends in fact on the set of active regulators, denoted in the sequel by ω_i.

6.3 Constraints on delays

More generally, it is possible to construct constraints on delays in order the system to follow a given sequence of domains. The principle of the construction of these constraints relies on the enumeration of constraints due to paths of length 2: $\mu_0 \to \mu_1 \to \mu_2$. For a longer path, the constraint is the conjunction of constraints due to each sub-path of length 2.

We describe here only one situation among twelve. Let us consider the path $\mu_0 \to \mu_1 \to \mu_2$ where the first (resp. second) transition is due to a

qualitative increasing of variable i_0 (resp. i_1). Let us suppose moreover that the vector $(c_i)_{i\in V}$ represents the clocks when entering into μ_1 and that there exists in μ_1 a variable i'_1 which can also increase. In order to allow the global path $\mu_0 \to \mu_1 \to \mu_2$, the following relation has to be satisfied:

$$(d^+_{i_1}(\mu_1) - c_{i_1}) < (d^+_{i'_1}(\mu_1) - c_{i'_1})$$

The twelve cases are exhaustively treated in [15].

6.4 Construction of constraints on FFL with auto-regulation

Let us consider the path allowing b to be switched-on before c. The sequence of domains is $(1,0,0) \to (2,0,0) \to (3,0,0) \to (3,1,0) \to (3,1,1) \to (3,0,1)$.

1. From $(1,0,0)$, there exists a unique successor domain: $(2,0,0)$. No constraint.

2. From $(2,0,0)$, there exists two possible successors: $(3,0,0)$ or $(2,0,1)$. Then, considering that clocks are reset to 0 when entering into $(2,0,0)$, we have:

$$d^+_a((2,0,0)) < d^+_c((2,0,0)) \qquad \text{see, Figure 16-left}$$

3. From $(3,0,0)$, it is possible to reach either $(3,1,0)$ or $(3,0,1)$. Then we have

$$d^+_b((3,0,0)) < d^+_c((3,0,0)) - d^+_a((2,0,0)) \qquad \text{see, Figure 16-right}$$

since during the crossing of the domain $(2,0,0)$, c has begun to increase.

4. From $(3,1,0)$ (resp. $(3,1,1)$), there exists a unique successor: $(3,1,1)$ (resp. $(3,0,1)$). No constraint.

Let us just remark that the delays deduced from the PLDE system of Figure 11-left (see equations 3, 5 and 6) does satisfy the previous contraints, whereas the delays deduced from the PLDE system of Figure 11-right does not satisfy them. Indeed taking into account parameters values of Figure 11-right, the analytic expression 5 gives $d^+_b((3,0,0)) = 2.77$ whereas expression 6 gives $d^+_c((3,0,0)) = 0.84$.
In a similar way, it is possible to build a set of constraints on delays which leads to trajectories along which the variable b is not switched on because of the fast increasing of c. The delays deduced from the PLDE system of Figure 11-left does not satisfy the contraints, whereas the delays deduced from the PLDE system of Figure 11-right does satisfy them.

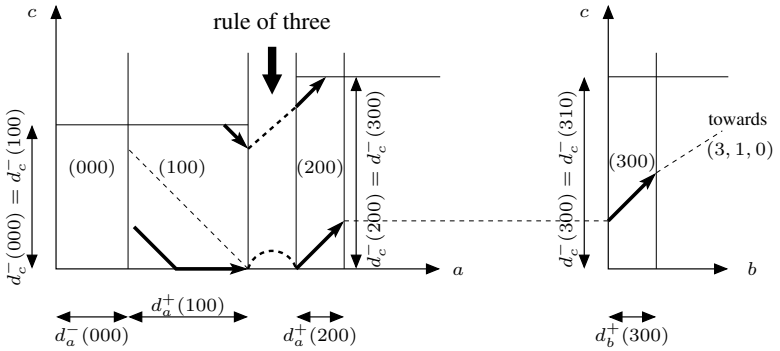


Figure 16: Illustration of the construction of constraints

6.5 A feedforward loop controled by a negative loop

We consider the boolean model, see Figure 17, of the feedforward loop controled by a negative loop which has been presented in Section 3.4: in order to build a boolean model thresholds of a on a', b and c are considered as equal. We would like to construct an hybrid system based on this boolean model

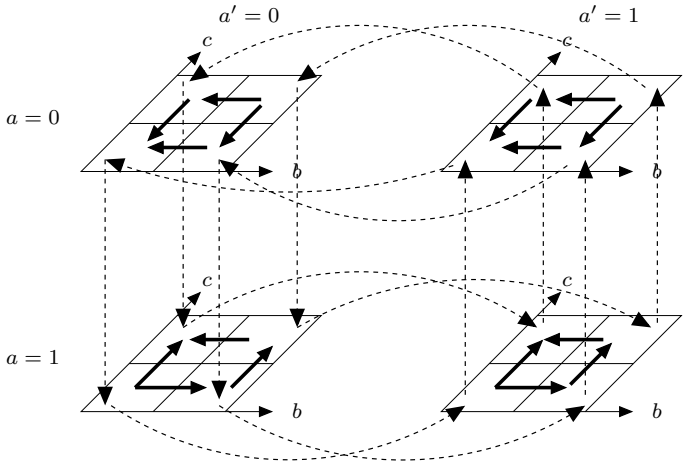


Figure 17: Boolean model of the FFL controled by a negative loop.

whose trajectories pass through the following sequence of domains:

$$(1000 \to 1100 \to 0100 \to 0000 \to) 1000 \to 1100 \to$$
$$(1110 \to 0110 \to 0010 \to 1010 \to)^2 1110 \to 1111$$

This path expresses that more than one period of oscillation of a and a' are mandatory to imply the qualitative increasing of b. Two other periods are

necessary to allow the qualitative activation of c which will be responsible of the degradation of b.

The expression of the corresponding constraints is very unreadable but is satisfiable. Figure 18 shows a set of values of parameters leading to the considered path.

$d^+_{a,0,\{a'\}}$	= 4	$d^+_{a',0,\{a\}}$	= 4
$d^+_{a,0,\emptyset}$	= 1	$d^+_{a',0,\emptyset}$	= 1
$d^-_{a,0,\emptyset}$	= 1	$d^-_{a',0,\emptyset}$	= 1
$d^+_{a,1,\{a'\}}$	= 1	$d^+_{a',1,\{a\}}$	= 1
$d^-_{a,1,\{a'\}}$	= 1	$d^-_{a',1,\{a\}}$	= 1
$d^-_{a,1,\emptyset}$	= 4	$d^-_{a',1,\emptyset}$	= 4
$d^+_{b,0,\{a,c\}}$	= 7	$d^+_{c,0,\{a\}}$	= 10
$d^+_{b,0,\emptyset}$	= 2	$d^+_{c,0,\emptyset}$	= 4
$d^-_{b,0,\emptyset}$	= 2	$d^-_{c,0,\emptyset}$	= 16
$d^+_{b,1,\{a,c\}}$	= 2	$d^+_{c,1,\{a\}}$	= 4
$d^-_{b,1,\{a,c\}}$	= 2	$d^-_{c,1,\{a\}}$	= 4
$d^-_{b,1,\emptyset}$	= 80	$d^-_{c,1,\emptyset}$	= 10

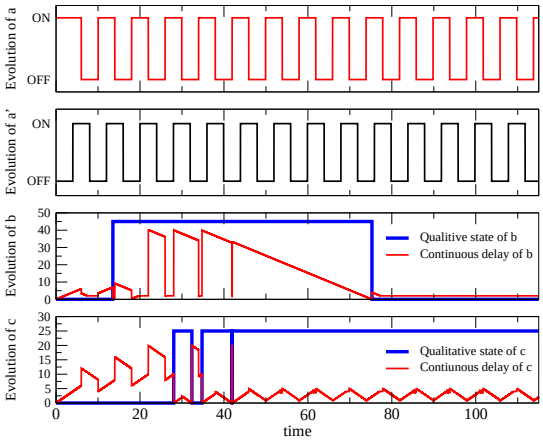


Figure 18: Simulation of a hybrid model for the FFL controled by a negative loop. The qualitative part of the initial state is $(1, 0, 0, 0)$, and its delay's part is $(2., 0., 0., 0.)$. Delays are denoted by the involved variable, its qualitative level and by the set of its effective regulators, see the description of the hybrid model.

7 Identification issues with delays

Let us remind that the cornerstone of the modelling activity is the parameter identification step. For the construction of a hybrid model based on the discrete modelling framework of R. Thomas, one has to identify both discrete parameters $K_{...}$ and delays $d^+_{...}$ and $d^-_{...}$ which correspond to approximations of times mandatory to pass through the involved domain.

In order to determine the values of delays, the modeller is going to rely on the measured elapsed time during experiments between two particular states. This global measured time has to be equal to the sum of delays of visited qualitative states. Thus, building the constraints associated with the measured elapsed time requires to know the sequence of visited qualitative states.

The parameter identification step can then be split into two subparts:

- To identify discrete parameters $K_{...}$ of the underlying discrete model. This step can be automated using model-checking or other formal methods (see sections 2.5, 2.6 and also [21, 13]).

- To identify the delays of the hybrid model. Here the built constraints on delays express relationships between a real number (the measured time) and a combination of delays. The resolutness of constraints on delays is unavoidable.

Let us mention some other approaches which do not consider a continuous time. The first way consists in discretizing the time in order to remain in a completely discrete modelling process. The underlying idea is to construct an approximation so fine as necessary as in the integral calculus. The second way consists in focusing on the duality between probabilistics approaches and models based on delays: greater the probability to fire a transition towards a particular qualitative state, smaller the associated delay. Thus all the scientific corpus of Markov chains can be useful to evaluate the probabilities of the model.

Conclusion

We have shown that different modelling frameworks for gene regulatory networks have been introduced. For all of them, the parameters have to be valuated. Fortunately this parameter identification step can be computer-aided when the modelling framework is formal and when it makes use of formal tools from computer science: for the purely discrete approach of R. Thomas, model checking, constraint programming or symbolic execution have been used to automate this stage. In order to take into consideration elapsed time and delays, it would be interesting to develop a tool that would take as inputs a PLDE system and a set of observed trajectories (and associated elapsed times) and that would give all possible valuations for parameters. Unfortunatelly such a tool is not conceivable for PLDE models. Thus, formal hybrid modellings seem to be the best candidates in order to fill up the gap between purely discrete models for which the parameter identification step can be automated and the differential models.

With such hybrid frameworks, systems biology should take advantage of the whole corpus of formal methods from computer science which opens a large horizon of research perspectives. Let us mention for instance,

- Algorithms that compute the set of parameter valuations that are compatible with reachability properties or with a known qualitative path.
- Continuous-time temporal logics adapted to the specificities of the biological application domain, and then model checking algorithms to confront a real-time temporal property to a hybrid model.
- Formal or symbolic languages to describe transition paths, taking into account populations of networks whose states are not synchronized.

- Since our aim is also to link modelling to experiments, tools to extract from the considered hybrid model several experiments which are able to refute the candidate models.

Indeed, the hybrid modellings are not the ultimate aim, they are only a guide for predictions that, in turn, suggest biological experiments whose success will be *in fine* the discriminent criterion. Thus a hybrid modelling framework will be largely adopted only if it is able to help biology toward comprehending the biological processes through the ability of the hybrid framework to propose experiments or through its capability of refuting hypotheses. Hybrid approaches could constitute a trade-off between expressiveness and computational tractability.

Acknowledgements

In a way, the biologist Janine Guespin is our ”muse”, who motivates by her enthusiasm a lot of our works. Even if she is not a co-author of this chapter, we thought of her when writing this chapter, as our typical target reader. This work could not have succeeded without the help and support of Olivier Roux and Jonathan Fromentin who developped with us the hybrid modelling framework. We would like to thank Adrien Richard and Laurent Trilling for their comments at different stages of this work. Finally, we are very grateful to Alexander Bockmayr, Heike Siebert for very deep discussions on how to introduce delays into discrete models of gene regulatory networks.

References

[1] J. Ahmad, G. Bernot, J.-P. Comet, D. Lime, and O. Roux. Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs*, 3(4):231–251, 2006 (Cover Date: Nov. 2007).

[2] J. Ahmad, O. Roux, G. Bernot, J.-P. Comet, and A. Richard. Analysing formal models of genetic regulatory networks with delays: Applications to lambda phage and T-cell activation systems. *Int. J. Bioinformatics Research and Applications*, 4(3):240–262, 2008.

[3] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.

[4] R. Alur, C. Courcoubetis, N. Halbwachs, T. A. Henzinger, P.-H. Ho, X. Nicollin, A. Olivero, J. Sifakis, and S. Yovine. The algorithmic analysis of hybrid systems. *T.C.S.*, 138(1):3–34, 1995.

[5] R. Alur, C. Courcoubetis, T. A. Henzinger, and P.-H. Ho. Hybrid automata: An algorithmic approach to the specification and verification of hybrid systems. In *Hybrid Systems*, pages 209–229, 1992.

[6] J. Bengtsson and W. Yi. Timed automata: semantics, algorithms and tools. In *Lecture Notes on Concurrency and Petri Nets*, number 3098 in LNCS, pages 87–124. Springer, 2008, See also http://www.uppaal.com.

[7] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *J.T.B.*, 229(3):339–347, 2004.

[8] F. Cassez and K. G. Larsen. The impressive power of stopwatches. In *CONCUR'00*, pages 138–152, London, UK, 2000. Springer-Verlag.

[9] A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV Version 2: An OpenSource Tool for Symbolic Model Checking. In *CAV 2002*, volume 2404 of *LNCS*, Copenhagen, Denmark, July 2002. Springer.

[10] F. Corblin, S. Tripodi, E. Fanchon, D. Ropers, and L. Trilling. A declarative constraint-based method for analysing discrete genetic regulatory networks. *Biosystems*, 98:91–104, 2009.

[11] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselman. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull. Math. Biology*, 2004.

[12] E.A. Emerson and J.Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. In *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing, San Francisco, California*, pages 169–180, 5-7 May 1982.

[13] E. Fanchon, F. Corblin, L. Trilling, B. Hermant, and D. Gulino. Modeling the molecular network controlling adhesion between human endothelial cells: Inference and simulation using constraint logic programming. In *CMSB 2004*, pages 104–118, 2005.

[14] A.F. Filippov. *Differential equations with discontinuous righthand sides*. Kluwer Academic Publishers, 1988.

[15] J. Fromentin, Bernot G., J.-P. Comet, and O. Roux. Hybrid modeling of gene regulatory networks: a combination of qualitative modeling with delays. *not yet published*, 2010.

[16] T.-A. Henzinger, P.-H. Ho, and H. Wong-Toi. HYTECH: A model checker for hybrid systems. *International Journal on Software Tools for Technology Transfer*, 1997.

[17] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press, 2000.

[18] H. Kitano. Biological robustness. *Nat. Rev. Genet.*, 5:826–837, 2004.

[19] H. Kitano. Towards a theory of biological robustness. *Mol. Syst. Biol.*, 3:137, 2007.

[20] S. Mangan and U. Alon. The structure and function of the feedforward loop network motif. *P.N.A.S. USA*, 100:11980–11985, 2003.

[21] D. Mateus, J.-P. Comet, J.-P. Gallois, and P. Le Gall. *Proc. of the Evry Spring school on Modelling and simulation of biological processes in the context of genomics*, chapter Inferring parameters of genetic regulatory networks with symbolic formal methods, pages 51–70. EDP Sci., 2007.

[22] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[23] H. Siebert and A. Bockmayr. Temporal constraints in the logical analysis of regulatory networks. *T.C.S.*, 391(3):258–275, 2008.

[24] E.H Snoussi. Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach. *Dynamics and stability of Systems*, 4:189–207, 1989.

[25] A.P. Thomas, A. Henzinger, P.W. Kopke, and P. Varaiya. What's decidable about hybrid automata. Technical report, EECS Department, University of California, Berkeley, 1998.

[26] R. Thomas. Regulatory networks seen as asynchronous automata : A logical description. *J.T.B.*, 153:1–23, 1991.

[27] R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, 1990.

[28] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. I. & II. *Chaos*, 11:170–195, 2001.

Hypothesis: the cytoskeleton is a metabolic sensor

Vic Norris[1], Patrick Amar[2], Guillaume Legent[1], Camille Ripoll[1], Michel Thellier[1] and Judit Ovádi[3]

[1] Laboratoire d'Assemblages moléculaires: modélisation et imagerie SIMS, Faculté des Sciences de l'Université de Rouen, F-76821 Mont Saint Aignan Cedex, France & DYCOEC, CNRS (GDR 2984), CORIA, Université de Rouen, France
[2] Laboratoire de Recherche en Informatique, Université Paris-Sud & UMR CNRS 8623, F-91405 Orsay cedex, France
[3] Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1113 Budapest, Hungary

Abstract

In the cytoskeletal sensor hypothesis proposed here, the cytoskeleton senses and integrates the general metabolic activity of the cell. This activity depends on the binding to the cytoskeleton of enzymes and, depending on the nature of the enzyme, this binding may occur if the enzyme is either active or inactive but not both. This enzyme-binding is further proposed to stabilise microtubules and microfilaments and to alter rates of GTP and ATP hydrolysis and their levels. These physical and chemical effects would have major consequences on cell shape, dynamics and cell cycle progression. Evidence consistent with the hypothesis is presented in the case of glycolysis and testable predictions are made.

1 Introduction

Cells face the enormous challenge of generating a single phenotype that must be coherent with a myriad internal and external conditions from hundreds of thousands - if not millions - of different constituents. Ensuring this coherence entails sensing and integrating a wide diversity of chemical and physical information so as to converge onto a few outputs. These outputs must affect many of the systems and hence must be extremely well-connected. Just how cells achieve this is far from clear.

One evident possibility is that metabolism and signalling are tightly linked. The concept of "functioning-dependent structure" (FDS) was developed to describe either those structures that only form when their constituents are performing a task (and that disappear when that these constituents cease performing the task) or the inverse, namely those structures that only form when the constituents are *not* performing their task [32]. A metabolic FDS comprises

enzymes that, for example, assemble into the higher order structure only when these enzymes are catalysing their reactions; in the absence of substrate these enzymes are therefore free [19]. Modeling the behaviour of such enzymes has revealed that they may be able to generate waves of metabolites and hence play a role in signaling [31]. A related concept is that of "ambiquitous" enzymes which can occupy two different positions in the cell, for example, free or associated with the cytoskeleton [18]. Here we bring together the concepts of FDS and ambiquity to explore the possibility that metabolism and signalling are linked via enzyme association with the cytoskeleton. We propose that a functioning-dependent co-assembly of metabolic enzymes plus cytoskeleton - a type of *enzoskeleton* [20] - could help solve the problem of generating a coherent phenotype.

2 The cytoskeletal sensor hypothesis

The binding to microtubules and to actin microfilaments of enzymes responsible for catalysing different metabolic pathways allows the cytoskeleton to sense and integrate metabolic activity. The transduction of this information occurs via alterations in the physical stability of cytoskeletal filaments and in the rates of hydrolysis of GTP and ATP.

2.1 The mechanism

1. Enzymes stabilise the cytoskeleton by binding to it. Certain enzymes may bind to the cytoskeleton when active, that is, catalysing their reactions, whilst others may only bind when inactive. The entire cytoskeletal hyperstructure or rather *enzoskeletal* hyperstructure is therefore an FDS.

2. Since the cytoskeleton and GTP and ATP are central players in cell structure and in most reactions, alterations in cytoskeletal dynamics due to enzyme binding and differential rates of GTP and ATP hydrolysis (and consequent levels) then alter cell structure and dynamics.

2.2 The evidence

Over a hundred proteins, including many involved in metabolism, change their distribution in the yeast, Saccharomyces cerevisiae, in response to altered metabolic conditions, moreover, association and dissociation of enzyme foci can be controlled by availability of specific metabolites, leading to the suggestion that metabolite-specific, reversible protein assemblies are common [17]. The actin cytoskeleton in S. cerevisiae also undergoes a major change as cells go from a quiescent state to growth. In the quiescent state, actin is in the form

of immobile bodies; on resumption of metabolic activity following refeeding, actin again forms the dynamic network [25].

2.2.1 Enzyme-binding to cytoskeleton

An extensive body of literature attests to the interactions of metabolic enzymes with microfilaments of actin and with microtubules. Interactions with microtubular proteins have been observed for the glycolytic kinases, hexokinase (HK), phosphofructokinase (PFK), and pyruvate kinase (PK) as well as aldolase (for references see [22]). These interactions between tubulin and metabolic enzymes lead to the formation of distinct hyperstructures (Figure 1 in [22]). Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) also binds to MT to cause MT bundling. A wide variety of enzymes, including translation factors, RNA-binding proteins, signalling proteins and metabolic enzymes - also interact with microtubules in plant cells [3]. Interactions between enzymes and MTs can be extended to interactions between MT motor proteins and enzymes. For example, the likely colocation of glucose-6-phosphate dehydrogenase with dynein has led to the proposal that microtubule motor proteins participate in hexose monophosphate shunt enzyme transport within leukocytes [11].

Interaction of glycolytic enzymes with actin microfilaments led, in the case of studies on muscle, to the conclusion that, in general, "actin binds enzymes". Even filamentous actin from yeast binds enzymes, albeit more weakly and such microfilaments bind aldolase and GAPDH [34].

2.2.2 Functioning-dependent binding to the cytoskeleton

The essence of our interpretation of the following results is that if the binding of an enzyme to the cytoskeleton increases the probability of catalysis (via for example its affinity for its substrate), reciprocally, an enzyme that is active in catalysis might well have a higher probability of associating with the cytoskeleton. In other words, if being bound to a cytoskeletal filament confers a conformation on an enzyme that allows it to bind its substrate then the activation of the free enzyme by substrate might promote the binding to this enzyme to the filament.

The binding of glycolytic enzymes to microtubules alters the catalytic and regulatory properties of these enzymes (see Table 1 in [22]). HK activity is increased by binding microtubules (resulting in enhanced glycolytic flux in brain tissue) but this does not influence their dynamics and structure. PFK activity is decreased by binding MTs (induced by enzyme dissociation of the tetrameric enzyme) and this periodically cross-links microtubules. PK activity

is unaffected by binding but this impedes MT assembly. The binding of the individual enzymes to MTs is influenced by enzyme-enzyme interactions. Formation of an aldolase - PFK complex prevents the association of PFK to MT or, put differently, results in PFK's detachment from the MT. Within this complex, PFK is stable and maintains its catalytic activity, the allosteric property is, however, abolished [22].

Results on PFK, GAPDH and aldolase interaction with microfilaments can also be interpreted as consistent with such interaction depending on the state of the enzyme. Binding to filamentous actin is known to activate PFK. Recently, it has been shown that insulin signaling increases the association of PFK with actin filaments and it was suggested that this association plays a role in the stimulation of glycolysis by insulin [23]. In serum-depleted cells, the cytoplasmic GAPDH is colocalised with actin stress fibres whereas in the presence of serum, this enzyme is distributed homogeneously [26]. In quiescent cells, aldolase is colocalised along stress fibres whereas in motile cells it is behind the ruffles at the leading edge of the cell [35].

2.3 Protein binding stabilises cytoskeleton to help determine phenotype

There is abundant evidence that proteins binding to the actin and tubulin cytoskeletons alter their dynamics. In the case of actin, the unregulated polymerisation of actin filaments is inhibited in cells by actin monomer-binding proteins such as profilin and Tbeta4 [4]. Nucleators of actin polymerisation include the Arp2/3 complex and its large family of nucleation-promoting factors (NPFs), formins, Spire, Cobl, VopL/VopF, TARP and Lmod. These proteins control the time and location for polymerisation and influence the structures of the actin networks. Coronin, an important protein in actin dynamics, changes its activity depending on the nucleotidic state of actin [6]. IQGAPs are actin-binding proteins that transmit extracellular signals to the actin network so as to influence mitogenic, morphological and migratory cell behaviour [16, 37]. In the case of tubulin, microtubule-associated proteins, such as tau and tubulin polymerisation-promoting protein (TPPP), promote MT assembly and stabilise MT networks with phosphorylation regulating these functions [10, 38]. For example, process extension in oligodendrocytes during differentiation is correlated with the increase in TPPP/p25 levels whilst a reduction in process extension is correlated with a decrease in TPPP/p25 levels; TPPP/p25, however, is not only co-localised with the microtubule network but is also found at the plasma membrane (where it is thought to mediate interactions between myelin basic protein and tubulin), consistent with TPPP/p25 possessing additional properties and specific roles depending on its location [36].

What then of the effects of metabolic enzymes binding to the cytoskeleton? It has been proposed that the specific cytoskeletal association of aldolase could play a structural role in cytoplasm and could contribute to metabolic regulation, metabolic compartmentation, and/or cell motility [35]; these authors also proposed that functional duality may be a widespread feature among cytosolic enzymes. It has also been proposed that the distribution of PFK activity could play a role in the metabolic regulation of breast cancer [5]. The importance of the marriage between metabolic enzymes and the actin cytoskeleton was shown in cells in which changes in the actin cytoskeleton were accompanied by changes in energy metabolism and it was proposed that an increased surface area for association between actin filaments and glycolytic enzymes enhances enzyme activity so as to supply more ATP [2]. An isotype of PK, pyruvate kinase M2, is a major regulator of the glycolytic flux in tumor cells and it has been suggested that M2-PK is a metabolic sensor which regulates cell proliferation, cell growth and apoptotic cell death in a glucose supply-dependent manner [27].

2.4 Testing the hypothesis

The following corollaries of the hypothesis could readily be tested experimentally or by simulation and modelling:

1. The numerous isotypes of actin and tubulin differ in their capacity to bind enzymes and, reciprocally, the isotypes of enzymes differ in their affinities for microfilaments and microtubules. This could be done in vitro or indeed in vivo using, for example, different fluorescent tags to distinguish between the behaviour of two different isotypes in the same cell.
2. There is a relationship between the sites on tubulin and actin that undergo post-translational modification and interaction with metabolic enzymes. This could be done as in 1/ and, in addition, use might be made of mutants.
3. The changes in rates of ATP and GTP hydrolysis that result from the stabilisation/destabilisation of cytoskeletal filaments are sufficient to alter the levels (and ratios) of ATP and GTP locally and perhaps globally in the cell and thereby exert important effects on the phenotype. Measurement (and simulation, see 4. below) of the intracellular levels of ATP and GTP under different conditions of cytoskeletal dynamics (and hence of nucleotide hydrolysis) should reveal whether the contribution of cytoskeletal dynamics to intracellular levels of ATP and GTP is sufficient to significantly affect local and perhaps global levels and ratios of these and related nucleotides.

4. Non-homogeneous spatio-temporal distributions of metabolites with a potential for signalling (as well as the distribution of the enzymes themselves) should be generated by functioning-dependent association or dissociation of enzymes from cytoskeletal filaments. Such distributions could be revealed by simulation with stochastic automata such as HSIM [1].

3 Discussion

The cytoskeletal sensor hypothesis is an attempt to answer the question of how cells sense and integrate a wide diversity of chemical and physical information so as to converge onto a few outputs and generate a coherent phenotype. This hypothesis is in itself insufficient but it can be combined with other, complementary, hypotheses to give an integrated picture of cell functioning.

One of these hypotheses is based on the possibility of ion condensation on the cytoskeleton [24]. Positive counterions such as potassium, magnesium, calcium and polyamines can condense onto negatively charged linear polymers [21]. Such condensation leads to the counterions being delocalised and diffusing in the near region in intimate contact with the polymer or other surface [15]. Condensation occurs at a critical value of the charge density of the polymer and resembles a phase transition in that it occurs in an abrupt fashion (for references see [24]). Since the activity of protein kinases and phosphatases can be modulated by ions, condensed ions on protein filaments might play a major role in the phosphorylation/dephosphorylation of a wide variety of protein filaments by filament-associated kinases/phosphatases. In this hypothesis, calcium condensation/decondensation on the macromolecular network creates coherent patterns of protein phosphorylation that transduce signals [24]. One of the attractive features of this hypothesis is that changes in temperature and in the tensional state of the macromolecular network cause changes in ion condensation on this network hence allowing it to integrate chemical and physical signals. Divalent ion condensation can also occur in vitro on MTs or actin microfilaments [14, 28, 29, 30, 8]. Since both actin and tubulin are negatively charged, the possibility exists that ion condensation on MT and microfilaments occurs in vivo. In the case of magnesium, condensation of this ion could provide a powerful basis for the activation of enzymes associated with the cytoskeleton whilst decondensation could activate enzymes free in the cytoplasm. Note too that cytoskeleton-associated enzymes could undergo conformational changes due to mechanotransduction by the cytoskeleton [12, 7] and that such changes can modulate catalysis [9]; This relationship may be two-way.

Another integrative hypothesis of relevance here is that of the cell as being a set of FDSs [33]. One central idea here is that enzymes that are free (i.e. unbound) are degraded preferentially. There is therefore spatial or configurational control over phenotype. This idea can be incorporated satisfactorily into the cytoskeletal sensing hypothesis. For example, an enzyme activated by substrate could bind to the cytoskeleton where it would be safe from proteases. Finally, in the cytoskeletal sensing hypothesis, the activity of numerous enzymes may be transduced by cytoskeletal dynamics into levels of ATP and GTP, two simple outputs with a myriad connections to cellular processes.

Acknowledgements

We thank the Epigenomics Project, Genopole, and the CNRS (DYCOEC, GDR 2984) for support.

References

[1] Amar, P., G. Legent, M. Thellier, C. Ripoll, G. Bernot, T. Nystrom, M. H. Saier, Jr. & V. Norris, (2008) A stochastic automaton shows how enzyme assemblies may contribute to metabolic efficiency. *BMC systems biology* **2**: 27.

[2] Bereiter-Hahn, J., C. Stubig & V. Heymann, (1995) Cell cycle-related changes in F-actin distribution are correlated with glycolytic activity. *Experimental cell research* **218**: 551-560.

[3] Chuong, S. D., A. G. Good, G. J. Taylor, M. C. Freeman, G. B. Moorhead & D. G. Muench, (2004) Large-scale identification of tubulin-binding proteins provides insight on subcellular trafficking, metabolic channeling, and signaling in plant cells. *Mol Cell Proteomics* **3**: 970-983.

[4] Dominguez, R., (2009) Actin filament nucleation and elongation factors–structure-function relationships. *Critical reviews in biochemistry and molecular biology* **44**: 351-366.

[5] El-Bacha, T., M. S. de Freitas & M. Sola-Penna, (2003) Cellular distribution of phosphofructokinase activity and implications to metabolic regulation in human breast cancer. *Molecular genetics and metabolism* **79**: 294-299.

[6] Gandhi, M., V. Achard, L. Blanchoin & B. L. Goode, (2009) Coronin switches roles in actin disassembly depending on the nucleotide state of actin. *Molecular cell* **34**: 364-374.

[7] Gardner, M. K., A. J. Hunt, H. V. Goodson & D. J. Odde, (2008) Microtubule assembly dynamics: new insights at the nanoscale. *Current opinion in cell biology* **20**: 64-70.

[8] Gartzke, J. & K. Lange, (2002) Cellular target of weak magnetic fields: ionic conduction along actin filaments of microvilli. *American journal of physiology* **283**: C1333-1346.

[9] Grinthal, A., I. Adamovic, B. Weiner, M. Karplus & N. Kleckner, (2010) PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 2467-2472.

[10] Hlavanda, E., E. Klement, E. Kokai, J. Kovacs, O. Vincze, N. Tokesi, F. Orosz, K. F. Medzihradszky, V. Dombradi & J. Ovadi, (2007) Phosphorylation blocks the activity of tubulin polymerization-promoting protein (TPPP): identification of sites targeted by different kinases. *The Journal of biological chemistry* **282**: 29531-29539.

[11] Huang, J. B., J. Espinoza, R. Romero & H. R. Petty, (2006) Apparent role of dynein in glucose-6-phosphate dehydrogenase trafficking in neutrophils from pregnant women. *Metabolism: clinical and experimental* **55**: 279-281.

[12] Ingber, D. E., (1997) Tensegrity: the architectural basis of cellular mechanotransduction. *Annual Review of Physiology* **59**: 575-599.

[13] Lehotzky, A., P. Lau, N. Tokesi, N. Muja, L. D. Hudson & J. Ovadi, (2010) Tubulin polymerization-promoting protein (TPPP/p25) is critical for oligodendrocyte differentiation. *Glia* **58**: 157-168.

[14] Lin, E. C. & H. F. Cantiello, (1993) A novel method to study the electrodynamic behavior of actin filaments. Evidence for cable-like properties of actin. *Biophysical journal* **65**: 1371-1378.

[15] Manning, G. S., (2007) Counterion condensation on charged spheres, cylinders, and planes. *J Phys Chem B* **111**: 8554-8559.

[16] Meyer, R. D., D. B. Sacks & N. Rahimi, (2008) IQGAP1-dependent signaling pathway regulates endothelial cell proliferation and angiogenesis. *PloS one* **3**: e3848.

[17] Narayanaswamy, R., M. Levy, M. Tsechansky, G. M. Stovall, J. D. O'Connell, J. Mirrielees, A. D. Ellington & E. M. Marcotte, (2009)

Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 10147-10152.

[18] Nemat-Gorgani, M. & J. E. Wilson, (1980) Ambiquitous behavior–a biological phenomenon of general significance? *Curr Top Cell Regul* **16**: 45-54.

[19] Norris, V., P. Gascuel, J. Guespin-Michel, C. Ripoll & M. H. Saier Jr., (1999) Metabolite-induced metabolons: the activation of transporter-enzyme complexes by substrate binding. *Molecular microbiology* **31**: 1592-1595.

[20] Norris, V., G. Turnock & D. Sigee, (1996) The Escherichia coli enzoskeleton. *Molecular microbiology* **19**: 197-204.

[21] Oosawa, F., (1971) *Polyelectrolytes.*, p. 1-160. Dekker, New York.

[22] Ovadi, J., F. Orosz & S. Hollan, (2004) Functional aspects of cellular microcompartmentation in the development of neurodegeneration: mutation induced aberrant protein-protein associations. *Mol Cell Biochem* **256-257**: 83-93.

[23] Real-Hohn, A., P. Zancan, D. Da Silva, E. R. Martins, L. T. Salgado, C. S. Mermelstein, A. M. Gomes & M. Sola-Penna, (2010) Filamentous actin and its associated binding proteins are the stimulatory site for 6-phosphofructo-1-kinase association within the membrane of human erythrocytes. *Biochimie*.

[24] Ripoll, C., V. Norris & M. Thellier, (2004) Ion condensation and signal transduction. *BioEssays* **26**: 549-557.

[25] Sagot, I., B. Pinson, B. Salin & B. Daignan-Fornier, (2006) Actin bodies in yeast quiescent cells: an immediately available actin reserve? *Molecular biology of the cell* **17**: 4645-4655.

[26] Schmitz, H. D. & J. Bereiter-Hahn, (2002) Glyceraldehyde-3-phosphate dehydrogenase associates with actin filaments in serum deprived NIH 3T3 cells only. *Cell biology international* **26**: 155-164.

[27] Spoden, G. A., U. Rostek, S. Lechner, M. Mitterberger, S. Mazurek & W. Zwerschke, (2009) Pyruvate kinase isoenzyme M2 is a glycolytic sensor differentially regulating cell proliferation, cell size and apoptotic cell death dependent on glucose supply. *Experimental cell research* **315**: 2765-2774.

[28] Tang, J. X. & P. A. Janmey, (1996) The polyelectrolyte nature of F-actin and the mechanism of actin bundle formation. *Journal of Biological Chemistry* **271**: 8556-8563.

[29] Tang, J. X., J. A. Kas, J. V. Shah & P. A. Janmey, (2001) Counterion-induced actin ring formation. *Eur Biophys J* **30**: 477-484.

[30] Tang, J. X., S. Wong, P. T. Tran & J. P.A., (1996) Counterion induced bundle formation of rodlike polyelectrolytes. *Ber. Bunsenges. Phys. Chem.* **100**: 796-806.

[31] Thellier, M., G. Legent, P. Amar, V. Norris & C. Ripoll, (2006) Steady-state kinetic behaviour of functioning-dependent structures. *The FEBS journal* **273**: 4287-4299.

[32] Thellier, M., G. Legent, V. Norris, C. Baron & C. Ripoll, (2004a) Introduction to the concept of "functioning-dependent structures" in living cells. *C:R: Biologies* **327**: 1017-1024.

[33] Thellier, M., V. Norris & C. Ripoll, (2004b) Modélisation du comportement cinétique des FDS (Functioning-dependent structures) dans les systèmes vivants. In: Comptes-Rendus de la 7ème rencontre du Non-Linéaire. Y. Pomeau & R. Ribotta (eds). Paris: Non-Linéaire Publications, Université de Paris-Sud, pp. 281-286.

[34] Waingeh, V. F., C. D. Gustafson, E. I. Kozliak, S. L. Lowe, H. R. Knull & K. A. Thomasson, (2006) Glycolytic enzyme interactions with yeast and skeletal muscle F-actin. *Biophysical journal* **90**: 1371-1384.

[35] Wang, J., D. R. Tolan & L. Pagliaro, (1997) Metabolic compartmentation in living cells: structural association of aldolase. *Experimental cell research* **237**: 445-451.

[36] Wenning, G. K., N. Stefanova, K. A. Jellinger, W. Poewe & M. G. Schlossmacher, (2008) Multiple system atrophy: a primary oligodendrogliopathy. *Annals of neurology* **64**: 239-246.

[37] White, C. D., M. D. Brown & D. B. Sacks, (2009) IQGAPs in cancer: a family of scaffold proteins underlying tumorigenesis. *FEBS letters* **583**: 1817-1824.

[38] Zheng, G. Q., X. M. Wang, Y. Wang & X. T. Wang, (2010) Tau as a potential novel therapeutic target in ischemic stroke. *Journal of cellular biochemistry* **109**: 26-29.

PART III TUTORIALS

Initiation to 3D Modeling and Visualization of biological processes

Damien Larivière[1], Eric Fourmentin[1], Ivan Junier[2], Joan Herisson[2], François Képès[2]

[1] Fondation Scientifique Fourmentin-Guilbert, 2 Avenue du Pavé Neuf, F-93160 Noisy-Le-Grand

[2] Epigenomics Project, Genopole® CNRS UPS 3201, UniverSud Paris, University of Evry, Genopole Campus 1 - Genavenir 6, 5 rue Henri Desbruères - F-91030 EVRY cedex

Abstract

Through 3D modeling and visualization, a biologist can experience space and crowding questions. This technique will help to generate new questions and new testable hypothesis. Another benefit is the creation of realistic images for scientific publications or funding purposes.

This (optional) practical work proposes an initiation to 3D modeling and visualization of biological environment containing numerous proteins and large portions of DNA.

Organization of the course We will first describe data (protein 3D structure, number, interaction, position, orientation) used to build a 3D scene and their origin (microscopy, simulation, empirical modeling).

We will then show how to transform these data into a 3D scene. The core of this process does not require to learn and use any modeler tools. The contribution of modeling tools to the 3D modeling process will be shortly addressed.

The biological viewer LifeExplorer (**`http://www.lifeexplorer.eu`**) will be used to navigate in and characterize the biological scenes created during the seminar. Cryo-EM reconstruction is not in the scope of the seminar.

Applications

- Rendering of the folding of a several thousands Kb DNA portion in bacterial nucleoid. The global compaction resulting from several states of folding will be compared as well as the volume accessible locally to various proteins. We will see how to generate a 3D scene from simulation data associating the long portion of DNA with compaction proteins and transcription enzymes. The rendering algorithm used here is well adapted to the study of viral DNA refolding into phage capside.
- Reconstruction of the FtsZ ring-like structure.
- Reconstruction of a crowded cytoplasm.

Pruning, pooling and limiting steps in metabolic networks

Ovidiu Radulescu[1,2], Alexander N. Gorban[3], Andrei Zinovyev[4]

[1] DIMNP - UMR 5235 CNRS/UM1/UM2, Pl. E. Bataillon, Université de Montpellier 2, Bat. 24 CP 107, 34095 Montpellier Cedex 5, France
[2] Team Symbiose, INRIA-IRISA, Campus de Beaulieu, F-35042 Rennes, France
[3] University of Leicester, LE1 7RH Leicester, UK
[4] Institut Curie, U900 INSERM/Curie/Mines ParisTech, 26 rue d'Ulm, F-75248 Paris, France

Abstract

Dynamics of metabolic systems can be modelled by systems of differential equations. Realistic models of metabolism allowing to integrate genome scale data should have very large size and thus face problems related to incompleteness of the information on their structure and parameters. We discuss how model reduction techniques that use qualitative information on the order of magnitude of parameters can be applied to simplify large models of differential equations.

1 Introduction

In spite of steady advance in the omic sciences (metabolomics, transcriptomics, proteomics), modelling of large biochemical networks, based on standard mathematical approaches, faces obstacles such as incompleteness of network description (structural and parametric) and lack of exact knowledge of kinetic parameters. In the particular case of the modelling of metabolic pathways, although genome scale reaction models are available for certain well studied organisms[2], complete and reliable information on the kinetic parameters of enzymatic reactions (Vm, Km) are not available for such very large models.

Constraint based approaches (such as flux balance analysis FBA [11]) circumvent these obstacles by using optimality principles and replacing the network by a set of stoichiometric constraints. FBA is well suited for global studies of perturbations of metabolism. Thermodynamics imposes constraints that can be dealt with within the same approach. FBA has been successfully applied to the global study of the metabolism of various organisms to identify the effects of gene knock-outs in various media, as well as for defining the concept of minimal supporting growth media. In spite of these successes FBA has two major drawbacks. It can not deal with time dynamics. Moreover it can not predict concentrations of metabolites (the predicted variables are the fluxes), that is a major defect when dealing with metabolomic studies.

Dynamic effects are particularly important in parmacokinetik when, depending on the dose time-scenario, the application of a drug could trigger or not compensatory mechanisms.

Dynamic modelling using differential equations was successfully used to study small or medium models (usually from ten to several tens of variables) for which complete sets of kinetic parameters can be measured or reverse-engineered. In order to study much larger models we could use model reduction techniques, which is a different way to tame complexity.

Model reduction is now common practice in combustion modelling where systems of thousands or tens of thousands of chemical reactions are reduced to much simpler sets of equations (see methods such as CSP, ILDM, invariant manifold [14, 12, 6, 5, 7, 1]). The feasibility of the reduction is guaranteed by generic properties of dissipative systems, that after a quick transition converge to a dynamics with a few degrees of freedom (invariant manifold [6, 5, 7, 1]). Applied to metabolic pathways such methods give a reduced description of dynamics in terms of synthetic variables that most of the time are difficult to interpret. This could be enough for numerical purposes such as stiffness elimination, but is not always appropriate for metabolic modelling. In this case we are looking for reduced variables that are easy to interpret and we need reduction methods that can cope with incomplete kinetic information. A recent extension of limitation theory [9, 15, 10] satisfies both these requirements. It can deal with both precise (exact values of constants) and qualitative (order of constants, such as much slower or much quicker) information. Also, the reduced models can be deduced from the initial model by simple constructions such as pruning and pooling.

2 Differential equations models

Continuous dynamical models of metabolism are conveniently represented as systems of differential equations. In such models, the state of the system is a vector $\boldsymbol{x} \in \mathbb{R}^n$ containing the concentrations of all metabolites. Each reaction (elementary step) in the model (indexed by an integer $i \in 1, \ldots, r$) is described by a stoichiometric vector $\boldsymbol{\nu}^i = \boldsymbol{\beta}^i - \boldsymbol{\alpha}^i$ ($\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ corresponds to the stoichiometries of products and reactants) and a rate R_i.

The rates (or fluxes) R_i, which are expressed in units of transformed mass per unit time, are functions of the concentrations. For reactions with no intermediate steps, occurring by random collisions of molecules, one can obtain the mass action law:

$$R(\boldsymbol{x}) = k_+ \prod_j x_j^{\alpha_j} - k_- \prod_j x_j^{\beta_j} \tag{1}$$

The set of reactions representing the step by step transformations of the metabolites is also called reaction mechanism or reaction network. The dynamics of a reaction network is given by the following system of differential equations:

$$\frac{d\boldsymbol{x}}{dt} = \sum_{i=1}^{r} R_i(\boldsymbol{x})\boldsymbol{\nu}^i \tag{2}$$

For instance, consider the Michaelis-Menten mechanism for enzymatic reactions $S + E \rightleftharpoons ES \longrightarrow E + P$.

For this model we have $\boldsymbol{x} = \begin{pmatrix} [S] \\ [E] \\ [ES] \\ [P] \end{pmatrix}$, $\boldsymbol{\nu}_1 = \begin{pmatrix} -1 \\ -1 \\ 1 \\ 0 \end{pmatrix}$,

$R_1 = k_1^+[S][E] - k_1^-[ES]$, $\boldsymbol{\nu}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \\ 0 \end{pmatrix}$, $R_2 = k_2[ES]$.

The system of differential equations reads:

$$\frac{d[S]}{dt} = -k_1^+[S][E] + k_1^-[ES] \tag{3}$$

$$\frac{d[E]}{dt} = -k_1^+[S][E] + k_1^-[ES] + k_2[ES] \tag{4}$$

$$\frac{d[ES]}{dt} = k_1^+[S][E] - k_1^-[ES] - k_2 ES \tag{5}$$

$$\frac{d[P]}{dt} = k_2[ES] \tag{6}$$

3 Traditional rate limiting step theory

In the IUPAC Compendium of Chemical Terminology (2007) one can find the following definition of limiting steps: “A rate-controlling (rate-determining or rate-limiting) step in a reaction occurring by a composite reaction sequence is an elementary reaction the rate constant for which exerts a strong effect - stronger than that of any other rate constant - on the overall rate.” Hans Krebs coined the term “pacemaker” for rate-limiting enzymes, that could play important role in targeting metabolism with drugs.

Although it is obvious from this definition that a rate-limiting step does not always exist (among the control functions generically there is a biggest

one, but this is not necessarily much bigger than all the others), biochemists tend to believe that each metabolic pathway has a unique limiting step even if most often do not agree on which one this is. On the other extreme, metabolic control theorists suggest that pathways rates depend to various degrees on rate constants of all the reactions, and thus limitation theory has limited utility. This theoretical prediction (relying on summation theorems for control coefficients [3]) seem to be verified experimentally. As cited by David Fell, a 3.5-fold increase of the amount of limiting enzyme phosphofructokinase in yeast have no significant effect on the anaerobic glycolytic flux [3]. True pacemaker enzymes allowing flux re-distributions are difficult to find.

However, for the notion of limiting step that is used in practice, there are important dynamical differences between systems without limiting step and systems with limiting step. The behavior of the later in terms of dynamics of intermediates and distribution of fluxes can be understood even if kinetic information is only partially quantitative. Finally, metabolic control and limitation theory can be unified in a common methodology. Limitation based model reduction can provide simpler models whose control coefficients can be more easily studied. Removing dominated (inessential) reactions allows to solve the problem of "sloppy sensitivities" identified in the context of gene networks, but also valid for regulation of metabolism.

4 Reducing linear networks with separated constants: pruning, glueing, and restoring

4.1 Linear networks

Linear reaction mechanisms include monomolecular networks or more generally first order networks.

The structure of monomolecular reaction networks can be completely defined by a simple digraph, in which vertices correspond to chemical species A_i, edges correspond to reactions $A_i \to A_j$ with rate constants $k_{ji} > 0$. In this case, the stoichiometric vector for the reaction (i, j) has -1, 1 in positions i, j, respectively and zeros elsewhere. The rate function is proportional to the concentration of the substrate $R_{ji}(\boldsymbol{x}) = k_{ji}x_i$.

The system of kinetic differential equations is

$$\frac{dx_i}{dt} = \sum_j k_{ij}x_j - (\sum_j k_{ji})x_i, \text{ or in matrix form } \frac{d\boldsymbol{x}}{dt} = \boldsymbol{K}\boldsymbol{x} \qquad (7)$$

where $\boldsymbol{K}$ is the kinetic matrix.

Monomolecular mechanisms are conservative, ie the total number of molecules is a constant of the dynamics $\sum_i x_i = const.$. This means that if a single

molecule is transformed by the mechanism, at each time there will be a single molecule somewhere in the mechanism.

The methods that we discuss here can be applied more generally to pseudo-conservative first order mechanisms.

First order reaction networks can contain reactions that are not monomolecular, such as $A \to A + B$, or $A \to B + C$. There is always a unique substrate and the rates proportional to the concentration of the substrate. These mechanisms are not conservative because they allow overall molecule production. Pseudo-conservative first order mechanism conserve total number of molecules of some species of interest that we call internal but can consume or produce external species. In pseudo-conservative first order mechanism $A \to A+B$ reactions are allowed, provided that B is external; similarly $A \to B+C$ reactions are allowed, provided that either B or C is external. Degradation reactions can be studied in this framework by considering a special component (sink), that collects degraded molecules.

Metabolic networks (or subnetworks) are rarely monomolecular or first order. However, when all substrates and cofactors are in excess, except for one, metabolic reactions can be also considered to be first order because in this case the rate is proportional to the concentration of the substrate that is not in excess. Though the general applicability of this method should not be taken for granted, linear formalism can provide new insights into metabolic network design.

4.2 Reduction algorithm for monomolecular networks

In [9, 15] we propose an algorithm to simplify monomolecular networks with total separation of the rate constants. Total separation of the constants means that either $k_I << k'_I$ or $k'_I << k_I$ for all $I = ij$, $I' = i'j'$. The algorithm, justified by estimates for the eigenvalues and eigenvectors (inspired, but not fully covered by Gershgorin theorem) of the kinetic matrix [8], consists of three stages:

I. Constructing of an auxiliary reaction network: pruning.

For each A_i branching node (substrate of several reactions) let us define κ_i as the maximal kinetic constant for reactions $A_i \to A_j$: $\kappa_i = \max_j\{k_{ji}\}$. For correspondent j we use the notation $\phi(i)$: $\phi(i) = \arg\max_j\{k_{ji}\}$.

An auxiliary reaction network $\mathcal{V}$ is the set of reactions obtained by keeping only $A_i \to A_{\phi(i)}$ with kinetic constants κ_i and discarding the other, slower reactions. Auxiliary networks have no branching, but they can have cycles and

confluences. The correspondent kinetic equation is

$$\dot{c}_i = -\kappa_i c_i + \sum_{\phi(j)=i} \kappa_j c_j, \tag{8}$$

If the auxiliary network contains no cycles, the algorithm stops here.

II gluing cycles and restoring cycle exit reactions

In general, the auxiliary network $\mathcal{V}$ has several cycles $C_1, C_2, ...$ with periods $\tau_1, \tau_2, ... > 1$.

These cycles will be "glued" into points and all nodes in the cycle C_i, will be replaced by a single vertex A^i. Also, some of the reactions that were pruned in the first part of the algorithm are restored with renormalized rate constants. Indeed, reaction exiting a cycle are needed to render the correct dynamics: without them, the total mass accumulates in the cycle, with them the mass can also slowly leave the cycle. Reactions $A \to B$ exiting from cycles ($A \in C_i$, $B \notin C_i$) are changed into $A^i \to B$ with the rate constant renormalization: let the cycle C^i be the following sequence of reactions $A_1 \to A_2 \to ...A_{\tau_i} \to A_1$, and the reaction rate constant for $A_i \to A_{i+1}$ is k_i (k_{τ_i} for $A_{\tau_i} \to A_1$). For the limiting (slowest) reaction of the cycle C_i we use notation $k_{\lim i}$. If $A = A_j$ and k is the rate reaction for $A \to B$, then the new reaction $A^i \to B$ has the rate constant $kk_{\lim i}/k_j$. This rate is obtained using quasi-stationary distribution for the cycle.

The new auxiliary network $\mathcal{V}^1$ is computed for the network of glued cycles. Then we decompose it into cycles, glue them, iterate until a acyclic network is obtained $\mathcal{V}^n$.

III Restoring cycles

The dynamics of species inside glued cycles is lost after the second part. A full multi-scale approximation (including relaxation inside cycles) can be obtained by restoration of cycles. This is done starting from the acyclic auxiliary network $\mathcal{V}^n$ back to $\mathcal{V}^1$ through the hierarchy of cycles. Each cycle is restored according to the following procedure:

For each glued cycle node A_i^m, node of $\mathcal{V}^m$,

- Recall its nodes $A_{i1}^{m-1} \to A_{i2}^{m-1} \to ...A_{i\tau_i}^{m-1} \to A_{i1}^{m-1}$; they form a cycle of length τ_i.
- Let us assume that the limiting step in A_i^m is $A_{i\tau_i}^{m-1} \to A_{i1}^{m-1}$
- Remove A_i^m from $\mathcal{V}^m$
- Add τ_i vertices $A_{i1}^{m-1}, A_{i2}^{m-1}, ...A_{i\tau_i}^{m-1}$ to $\mathcal{V}^m$
- Add to $\mathcal{V}^m$ reactions $A_{i1}^{m-1} \to A_{i2}^{m-1} \to ...A_{i\tau_i}^{m-1}$ (that are the cycle reactions without the limiting step) with correspondent constants from $\mathcal{V}^{m-1}$

- If there exists an outgoing reaction $A_i^m \rightarrow B$ in $\mathcal{V}^m$ then we substitute it by the reaction $A_{i\tau_i}^{m-1} \rightarrow B$ with the same constant, i.e. outgoing reactions $A_i^m \rightarrow ...$ are reattached to the beginning of the limiting steps

- If there exists an incoming reaction in the form $B \rightarrow A_i^m$, find its prototype in $\mathcal{V}^{m-1}$ and restore it in $\mathcal{V}^m$

- If in the initial $\mathcal{V}^m$ there existed a "between-cycles" reaction $A_i^m \rightarrow A_j^m$ then we find the prototype in $\mathcal{V}^{m-1}$, $A \rightarrow B$, and substitute the reaction by $A_{i\tau_i}^{m-1} \rightarrow B$ with the same constant, as for $A_i^m \rightarrow A_j^m$ (again, the beginning of the arrow is reattached to the head of the limiting step in A_i^m)

The result of the algorithm is a reduced network that has no cycles and no branchings. Some reactions necessarily disappear from the initial model in order to break cycles and eliminate branchings, so the global operation can be called pruning. Pruning expresses the domination relations between pathways. Simple in acyclic networks (the quicker branch dominates, the much slower branches are pruned), these relations can be quite intricate in the presence of cycles. Rate constants of some of the remaining reactions are changed into monomial functions of the initial constants.

For the reduced network the calculation of the dynamics is straightforward.

Solution of the homogeneous linear dynamic equations (7) are:

$$x(t) = \sum_{k=1}^{n} r^k (l^k, x(0)) \exp(\lambda_k t) \tag{9}$$

where λ_k, l^k, r^k are the eigenvalues, left and right eigenvectors of the kinetic matrix $\boldsymbol{K}$, respectively: $l^k \boldsymbol{K} = \lambda_k l^k$, $\boldsymbol{K} r^k = \lambda_k r^k$.

Computing eigenvalues and eigenvectors is straightforward for acyclic networks with no branching.

The eigenvalues are $\lambda_i = -\kappa_i$, one for each node in the network. If a node i is a sink (it has no successor) we consider that $\lambda_i = 0$.

Right eigenvectors r^i are obtained by recursion, in the forward direction along the reaction graph. One has $r_j^i = 0$ for $j < i$. Starting with the normalised value $r_i^i = 1$, the coordinates $r_{\phi^k(i)}^i$ $(k = 1, 2, ..)$ are obtained by:

$$r^i_{\phi^{k+1}(i)} = \frac{\kappa_{\phi^k(i)}}{\kappa_{\phi^{k+1}(i)} - \kappa_i} r^i_{\phi^k(i)} = \prod_{j=0}^{k} \frac{\kappa_{\phi^j(i)}}{\kappa_{\phi^{j+1}(i)} - \kappa_i} = \frac{\kappa_i}{\kappa_{\phi^{k+1}(i)} - \kappa_i} \prod_{j=1}^{k} \frac{\kappa_{\phi^j(i)}}{\kappa_{\phi^j(i)} - \kappa_i}. \tag{10}$$

Left eigenvectors are also obtained by recursion, but in the reverse direction. Thus, $l^i_j = 0$ for $j > i$. Starting with the normalised value $l^i_i = 1$, the coordinates l^i_j are obtained as:

$$l^i_j = \frac{\kappa_j}{\kappa_j - \kappa_i} l^i_{\phi(j)} = \prod_{k=0}^{q-1} \frac{\kappa_{\phi^k(j)}}{\kappa_{\phi^k(j)} - \kappa_i}. \tag{11}$$

In the case of fully separated systems, these expressions are significantly simplified and do not require knowledge of the exact values of κ_i. Thus, for the left eigenvectors $l^i_i = 1$ and, for $i \neq j$,

$$l^i_j = \begin{cases} 1, \text{ if } \phi^q(j) = i \text{ for some } q > 0 \text{ and } \kappa_{\phi^d(i)} > \kappa_i \text{ for all } d = 0, \dots q-1 \\ 0, \text{ else} \end{cases} \tag{12}$$

For the right eigenvectors we suppose that $\kappa_f = 0$ for a sink vertex A_f. Then $r^i_i = 1$ and

$$r^i_{\phi^k(j)} = \begin{cases} -1, \text{ if } \kappa_{\phi^k(i)} < \kappa_i \text{ and } \kappa_{\phi^m(i)} > \kappa_i \text{ for all } m = 1, \dots k-1 \\ 0, \text{ else} \end{cases} \tag{13}$$

A monomolecular network with totally separated constants have rate-limiting step. Supposing that the reduced network is a chain, the rate-limiting step is the slowest reaction in the chain. However, this is not always the slowest reaction of the initial network.

Broken cycle The simplest example illustrating this counterintuitive possibility is a cycle of reactions. Consider an isolated cycle with total separation. The reduced acyclic model is the chain obtained from the the cycle by removing its slowest constant. The rate limiting step in the chain is the second slowest constant of the cycle. A cycle with total separation behaves like a chain ensuring transport of the mass to the beginning of the slowest step.

Interrupted pathway The effect of pruning can also lead to pathway interruption. For instance let us consider the example in Fig. 1 below.

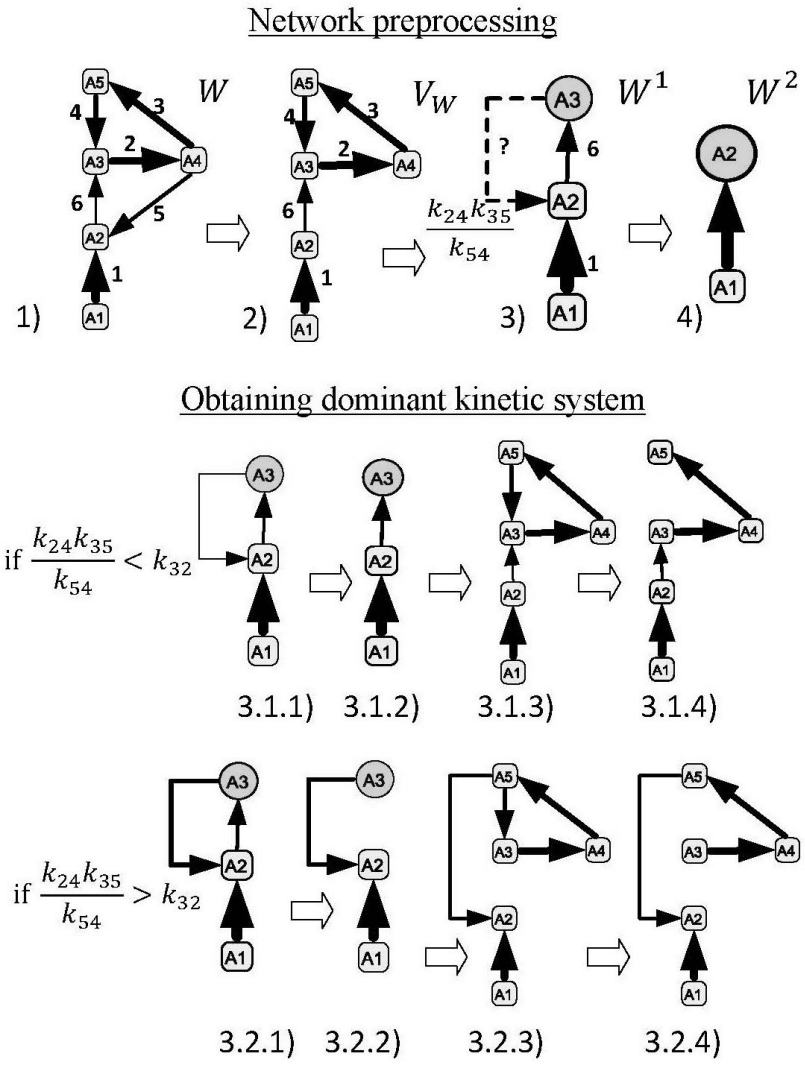


Figure 1: Example of calculation of the dominant approximations for a monomolecular reaction network with total separation of the constants (from [15]). The order of kinetics parameters is shown both by integer numbers (ranks) and the thickness of arrows (faster reactions are thicker).

The network preprocessing consists in pruning reaction $A_4 \rightarrow A_2$ because this is dominated by the much faster reaction $A_4 \rightarrow A_5$. The resulting auxiliary network has the cycle (A_3, A_4, A_5), that is glued to A_3 in the step 3). The reaction $A_4 \rightarrow A_2$ exiting the cycle is restored with renormalized $k'_{24} = k_{24}k_{35}/k_{54}$ constant in step 3). This produces a new cycle that is glued to A_2. Depending on the order relation between the renormalized constant k'_{24} of the exit reaction and the constant k_{32}, the limiting step of the glued cycle A_2 can change. After elimination of cycle limiting step and cycle restoration 3.1.3-4 or 3.2.3-4 there are two possible reduced neworks, both of them chains. In the case 3.1.4 the limiting step for the transformation of A_1 into A_5 is the reaction $A_2 \rightarrow A_3$, the slowest reaction in the initial mechanism, but in 3.2.4 the reduced mechanism no longer contains this transformation.

Futile cycles and switching Metabolic networks contain many cyclic structures. As discussed in [3] a futile cycle converts a metabolite into another and back. It produces no net change but dissipates energy. Among various potential roles of futile cycles (heat production, increased control coefficients) there is the possibility of switching the direction of the flux (see Fig.2 below).

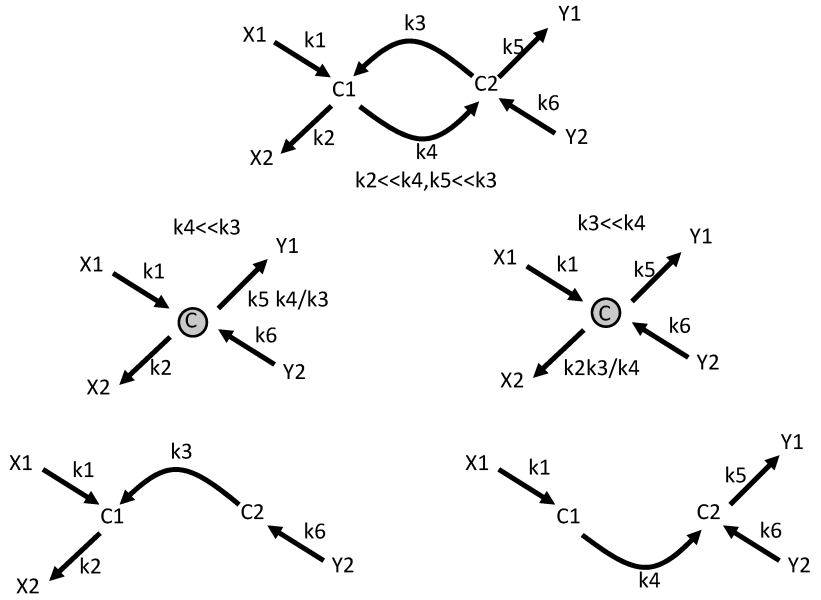


Figure 2: Futile cycle used as a metabolic switch. The cycling condition reads $k_2 << k_4, k_5 << k_3$.

5 Quasi-stationarity and quasi-equilibrium: pooling species and reactions

Quasi-stationarity and quasi-equilibrium are useful concepts that can be used for model reduction of rather general reaction mechanisms.

Quasi-equilibrium reactions are reversible reactions at thermodynamic equilibrium. Species involved in quasi-equilibrium reactions bear well defined algebraic relations between their concentrations. A mass action quasi-equilibrium reaction would imply:

$$\frac{\prod_j x_j^{\beta_j}}{\prod_j x_j^{\alpha_j}} = K_{eq} = \exp\left(-\frac{\Delta G}{RT}\right) \tag{14}$$

where R is the universal gas constant, T is the temperature, ΔG is the Helmholtz free energy change, $K_{eq} = k^+/k^-$.

Slightly more generally, quasiequilibrium approximation uses the assumption that a group of reactions is much faster than other and goes fast to its equilibrium. This can be studied by using singular perturbations [18, 19], by introducing a small positive parameter ϵ representing the ratio of timescales of slow and fast reactions. Then the dynamics reads:

$$\frac{d\boldsymbol{x}}{dt} = \sum_{s,slow} R_s \boldsymbol{\gamma}^s + \frac{1}{\epsilon} \sum_{f,fast} R_f \boldsymbol{\gamma}^f \tag{15}$$

To separate slow/fast variables, we have to study the spaces of linear conservation law of the initial system and of the fast subsystem:

$$\frac{d\boldsymbol{x}}{dt} = \frac{1}{\epsilon} \sum_{f,fast} R_f \boldsymbol{\gamma}^f \tag{16}$$

In general the system (15) can have several conservation laws. These are linear functions $b^1(\boldsymbol{x}), \ldots, b^m(\boldsymbol{x})$ of the concentrations that are constant in time. The conservation laws of the system (16) provide variables that are constant on the fast timescale. If they are also conserved by the full dynamics, the system has no slow variables (variables are either fast or constant). In this case, the dynamics of the fast variables is simply given by Eq.(16). Suppose now that the system (16) has some more conservation laws $b^{m+1}(\boldsymbol{x}), \ldots, b^{m+l}(\boldsymbol{x})$ that are not conserved by the full system (15). Then, these provide the slow variables of the system. The QE equation $\sum_{f,slow} R_f \boldsymbol{\gamma}^f = 0$ serves to compute fast variables as functions of the slow ones [10].

The quasisteady-state (QSS) assumption was invented in chemistry for description of systems with radicals or catalysts [20]. In the most usual version

[16], the species are split in two groups with concentration vectors $\boldsymbol{c}^s$ ("slow" or basic components) and $\boldsymbol{c}^f$ ("fast intermediates" or QSS species).

The small parameter ϵ used in singular perturbation theory is now the ratio of small concentrations of fast intermediates to the concentration of other species. After rescaling $\boldsymbol{c}^s$ and $\boldsymbol{c}^f$ to order one, the set of kinetic equations reads:

$$\frac{d\boldsymbol{c}^s}{dt} = \boldsymbol{W}^s(\boldsymbol{c}^s, \boldsymbol{c}^f) \tag{17}$$

$$\frac{d\boldsymbol{c}^f}{dt} = (1/\epsilon)\boldsymbol{W}^f(\boldsymbol{c}^s, \boldsymbol{c}^f) \tag{18}$$

where the functions $\boldsymbol{W}^s$, $\boldsymbol{W}^f$ and their derivatives are of order one ($0 < \epsilon << 1$).

The standard singular perturbation theory[18, 19] provides the QSS algebraic condition $\boldsymbol{W}^f(\boldsymbol{c}^s, \boldsymbol{c}^f) = 0$. These equations, together with additional balances for c^f (conservation laws) are enough to deduce the fast variables c^f as functions of the slow variables c^s and to eliminate them [20, 13, 15]. The slow dynamics is given by Eq.(17).

However, not all fast species are small concentration intermediates. The simplest such example is a fast irreversible cycle with a slow exit reaction. This example does not correspond to the traditional definition of quasi-equilibrium because it lacks reversibility and can not fulfill detailed balance. A singular perturbation analysis similar to the one for QE shows that the total mass of the cycle (this can be arbitrarily big) represents a slow variable, while each one of the concentrations of species inside the cycle are fast variables. The algebraic relations for fast species are those for QSS because they express the steady-state condition for the fast cycle.

The simplest illustration of these two approximation is provided by the Michaelis-Menten model for enzymatic reaction (see section 2).

One can have quasi-equilibrium if the first equation is fast: $k_1^{\pm} = \kappa^{\pm}/\epsilon$, where $\epsilon > 0$ is small. Then, the quantities conserved by the rapid reaction form two slow pool global variables, namely $C^s = [S] + [ES]$ and $b_E = [E] + [ES]$. Actually, b_E is conserved by all the reactions of the mechanism, so it is a kinetics constant $b_E = const.$. The algebraic quasi-equilibrium condition reads $k_1^+(C^s - [ES])(b_E - [ES]) = k_1^-[ES]$. This gives a dependence of the pool variable $[ES]$ on the pool global variables, namely $[ES] = c_{ES}(C^s, b_E)$. The final slow dynamics, obtained from Eq.6 and by adding term by term the Eqs.3,5, reads:

$$\dot{C^s} = -k_2 c_{ES}(C^s, b_E) \quad (19)$$
$$\dot{P} = k_2 c_{ES}(C^s, b_E) \quad (20)$$

The reduced QE mechanism is a single reaction transforming the pool C^s into the product P, with a rate law $R(C^s) = k_2 c_{ES}(C^s, b_E)$ (the constant b_E is the total quantity of enzyme).

The QSS approximation is obtained from the full mechanism when the enzyme is in much less quantity than the substrate $C^s >> b_E$. Under this condition, E and ES are fast intermediates (QSS species). b_E is conserved and constant like before. The slow variables are the concentrations of non-intermediate species, namely $[S], [P]$.

The QSS algebraic condition reads $k_1^+[S][E] = (k_1^- + k_2)[ES]$ which gives a dependence of the fast variables on the concentrations of the other species:

$$[ES] = c_{ES}([S], b_E) = k_1^+[S]b_E/(k_1^+[S] + k_1^- + k_2) \quad (21)$$

The dynamics of the external (non-intermediate) species reads:

$$[\dot{S}] = -k_1^+[S](b_E - c_{ES}([S])) + k_1^- c_{ES}([S]) = -k_2 c_{ES}([S], b_E) \quad (22)$$
$$[\dot{P}] = k_2 c_{ES}([S], b_E) \quad (23)$$

The reduced QSS mechanism is a single reaction transforming the substrate S into the product P, with a rate law $R_{MM}(C^s, b_E) = k_2 c_{ES}(C^s, b_E) = V_m[S]/([S]+K_m)$, where R_{MM} is the well known Michaelis-Menten rate law, $V_m = k_2 b_E$, $K_m = (k_1^- + k_2)/k_1^+$.

The single step of the reduced M-M mechanism can be seen as resulting from merging (considering them to be simultaneous) two steps $S + E \longrightarrow ES$ and $ES \longrightarrow E + P$ of the initial full mechanism. This merged step, or "pool of reactions", is a combination of reactions involving the two rapid species E and ES such that the resultant reaction does not change the concentrations of any of the fast species (these combinations conserve the fast species). Given a reaction mechanism and a set of fast species, there may be several reaction pools that preserve fast species (notice that for pool definition, reversible reactions are considered as two steps, one for each direction). There is only one such pool for the M-M mechanism.

The difference between the QSS and the QE in this example is obvious.

QE corresponds to pooling of species. For QE we first identify pools of species that are rapidly transformed one into the other by rapid QE reactions.

The number of pools is given by the number of conservation laws of the set of QE reactions that are not conservation laws of the full mechanism. Then, QE conditions allow to express the rates of reactions exiting the pools as functions of global variables of the pools (conservation laws of QE reactions). The reduced mechanism is made out of the pools, the remaining species and reactions with rate laws thus computed.

QSS corresponds to a pooling of reactions. A reaction pool (also called reaction route [20, 17]) is a linear combination with positive integer coefficients of reactions in the mechanism (reversible reactions counted twice, one reaction for each direction). We are interested in those pools (routes) that transform slow species into other slow species and conserve the intermediate fast species. In [15] we have also imposed a simplicity criterion for the pools, by choosing only simple sub-mechanisms. Simple sub-mechanisms are pools (routes) with a minimal number of reactions, transforming slow species without producing accumulation or depletion of the intermediate fast species. According to this definition, simple sub-mechanisms are elementary modes [4] of the set of reactions involving fast species. The QSS conditions and the internal balances are used to express the concentration of intermediate species and the rate laws of pooled reactions as functions of the concentrations of slow species [15].

Coming back to the previous section we would like to relate the reduction algorithm for monomolecular networks to the general concepts of QE and QSS.

Monomolecular networks with completely separated constants can not be considered to be at quasi-equilibrium, because they do not include reversible reactions (if both forward and reverse fluxes are allowed, then one of them dominates the other). Although quasi-equilibrium ideas have been used as an intermediate step of the reduction algorithm (gluing cycles), the reduced model is a acyclic graph with no pooling of species.

Monomolecular networks with completely separated constants can contain QSS species. These can be easily identified in the reduced model which is a chain or a set of chains with confluences. For instance, if the reduced model is a chain, the QSS species are those species that are consumed by fast reactions. In totally separated chains, QSS species concentrations can be set to zero (they are consumed by fast reactions). For instance, in the example shown in Fig.1, case 3.1.4 for timescales of the order of the inverse of the rate limiting step, three species A_1, A_3, A_4 are QSS. The reaction pool $(A_2 \longrightarrow A_3) + (A_3 \longrightarrow A_4) + (A_4 \longrightarrow A_5)$ gives the reduced mechanism $A_2 \longrightarrow A_5$ of constant rate k_{32}.

As suggested above, identification of QSS species, QE reactions and limiting steps is not easy in general. The QSS and QE nature of species and reactions as well as the limiting steps are global properties of the reaction mechanism that can not be easily obtained by comparing rate constants of individual reactions. Furthermore, the idea of "rapid reactions" can lead to very complex kinetic situation and should be used with care in the reduction of models.

Last but not least, we must emphasize an important difference between QSS and QE. Contrary to QE, QSS is a purely kinetic concept and has no relation to equilibrium thermodynamics (it does not have to obey detailed balance for instance). Thus, in the QSS situation, rate constants can not be related to thermodynamic potentials. This makes QE a simpler situation from the point of view of parameter identification.

6 Conclusion

Dynamics of metabolic networks can be studied by systems of differential equations. Large models with incomplete information are not suited for immediate analysis by traditional approaches and have to be simplified.

We have presented several model reduction techniques allowing to transform large reaction networks into simpler networks, whose dynamics can be readily studied. These techniques exploit the separation of the timescales of the complex networks. In the process of simplification, non-critical elements are removed from the models, and only essential elements are kept.

For monomolecular networks with total separation of the rate constants, we propose a reduction algorithm allowing to transform any such network into an acyclic network without branching, whose dynamics is computed analytically. The global transformation leading to simpler monomolecular networks can be defined as pruning. This transformation eliminates dominated reactions and computes a dominant subnetwork. The limiting step, easily identified on the reduced network, can be different from the slowest reaction of the full mechanism. Monomolecular models, though not always realistic, can teach us about design principles of large networks.

More general concepts such as quasi-equilibrium and quasi-steady state approximation can be applied to simplify non-linear as well as linear networks. We showed how these approximations can be related to pooling of species and of reactions.

Pooling of species and of reactions can also result from decompositions of the Jacobian (matrix defining the linearised dynamics) of nonlinear systems

of differential equations (2), once all rate constants are known. This method has been applied in [11] to analyse pooling of a fully parametrized glycolysis model. However, one would like to obtain the pools without knowing numerical values of all the parameters, using only the order relations between time scales and/or rate constants.

There is still much to do in this direction to propose simple general rules allowing for correct identification of limiting steps, QSS species and QE reactions. The next case to study will be the linear networks with partial separation, that could be approached by a combination of pooling and pruning. QSS and QE, combined with techniques for dominant solutions of algebraic equations represent a promising approach to the reduction of non-linear models (see [15, 13]).

References

[1] E. Chiavazzo, A.N. Gorban, and I.V. Karlin. Comparison of invariant manifolds for model reduction in chemical kinetics. *Communications in Computational Physics*, 2(5):964–992, 2007.

[2] A.M. Feist, C.S. Henry, J.L. Reed, M. Krummenacker, A.R. Joyce, P.D. Karp, L.J. Broadbelt, V. Hatzimanikatis, and B.Ø. Palsson. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3(1), 2007.

[3] D. Fell. *Understanding the control of metabolism*. Portland Press London, 1997.

[4] J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC bioinformatics*, 5(1):175, 2004.

[5] Alexander N. Gorban, Iliya V. Karlin, and Andrei Yu. Zinovyev. Constructive methods of invariant manifolds for kinetic problems. *Physics Reports*, 396(4-6):197 – 403, 2004.

[6] A.N. Gorban and I.V. Karlin. Method of invariant manifold for chemical kinetics. *Chemical Engineering Science*, 58(21):4751–4768, 2003.

[7] A.N. Gorban and I.V. Karlin. *Invariant manifolds for physical and chemical kinetics, Lect. Notes Phys. 660*. Springer Verlag, Berlin, Heidelberg, 2005.

[8] A.N. Gorban and O. Radulescu. Dynamical robustness of biological networks with hierarchical distribution of time scales. *IET Syst. Biol*, 1(4):238–246, 2007.

[9] AN Gorban and O. Radulescu. Dynamic and static limitation in reaction networks, revisited . In David West Guy B. Marin and Gregory S. Yablonsky, editors, *Advances in Chemical Engineering - Mathematics in Chemical Kinetics and Engineering*, volume 34 of *Advances in Chemical Engineering*, pages 103–173. Elsevier, 2008.

[10] AN Gorban, O. Radulescu, and AY Zinovyev. Asymptotology of chemical reaction networks. *Chemical Engineering Science*, 65:2310–2324, 2010.

[11] N. Jamshidi and B.Ø. Palsson. Formulating genome-scale kinetic models in the post-genome era. *Molecular Systems Biology*, 4:171, 2008.

[12] SH Lam and DA Goussis. The CSP method for simplifying kinetics. *International Journal of Chemical Kinetics*, 26(4):461–486, 1994.

[13] Mark Z. Lazman and Gregory S. Yablonsky. Chapter 2 overall reaction rate equation of single-route complex catalytic reaction in terms of hypergeometric series. In David West Guy B. Marin and Gregory S. Yablonsky, editors, *Advances in Chemical Engineering - Mathematics in Chemical Kinetics and Engineering*, volume 34 of *Advances in Chemical Engineering*, pages 47 – 102. Elsevier, 2008.

[14] U. Maas and S.B. Pope. Simplifying chemical kinetics- Intrinsic low-dimensional manifolds in composition space. *Combustion and Flame*, 88(3-4):239–264, 1992.

[15] O. Radulescu, A.N. Gorban, A. Zinovyev, and A. Lilienbaum. Robust simplifications of multiscale biochemical networks. *BMC systems biology*, 2(1):86, 2008.

[16] L.A. Segel and M. Slemrod. The quasi-steady-state assumption: a case study in perturbation. *SIAM Review*, 31(3):446–477, 1989.

[17] ON Temkin, A.V. Zeigarnik, and D. Bonchev. *Chemical Reaction Networks: a graph-theoretical approach*. CRC, Boca Raton, 1996.

[18] A. N. Tikhonov, A. B. Vasileva, and A. G. Sveshnikov. *Differential equations*. Springer, Berlin, 1985.

[19] W. Wasow. *Asymptotic Expansions for Ordinary Differential Equations*. Wiley, New York, 1965.

[20] GS Yablonskii, VI Bykov, AN Gorban, and VI Elokhin. *Kinetic models of catalytic reactions. Comprehensive Chemical Kinetics*, volume 32. Elsevier, Amsterdam, 1991.

PART IV POSTERS

Obliterating the phylogenetic bias in multiple sequence alignment.

Arianna Bertolino[1,*], Bryan Lunt[2,*], Peter Arndt[3], Martin Weigt[1]

[1]Institute for Scientific Interchange, Viale S. Severo 65, 10133 Turin, Italy
[2]UCDS, Department of Physics, University of California at San Diego
[3]Max Planck Institute for Molecular Genetics, Ihnestraβe 73, Berlin, Germany

[*] these authors contributed equally to this work

Abstract

The increasing availability of genomic data offers a new opportunity to search for specific signals of the coevolution of amminoacids in proteins. When the amino acid at one site changes, it can alter the selective forces associated with other sites, thus altering the set of mutations selectively admissible at those sites. The coevolution between two residues can be estimated thanks to the mutual information, an attractive metric that esplicitly measures the codependencies between two random variables. Using multiple sequence alignments to estimate the distribution of the amino acid in a site, mutual information quantifies how much uncertainty in the aminoacid state in one site can be removed by the knowledge at the amino acid state at another site.

The easiest way to estimate the nucleotide or amino acid distribution at one locus in a functional genomic region is to observe and count the respective state in a number of species. However these species did not evolve independently from each other and the frequency count is showing a phylogentic bias.

The main idea of our work is that this bias can be reduced by observing the same species after they evolved independently for a long time on evolutionary time scales where information about a common ancestral state is eventually lost. After this time the distribution of frequency counts will be the stationary distribution of the Markov process that governs the evolution of this site due to functional constraints.

A maximum likelihood approach enables us to reconstruct this Markov process for each site and infer the stationary state, the unbiased estimator (phylogenetic) for the state distribution. The performance of the algorithm depends on several parameters as the depth of the phylogeny, the number of states, the number of leaves, the topology of the phylogeny.

The effectivness of the algorithm has been tested for 2,4 and 20 states case on several ultrametric phylogenies with a coupled work of analytics and sim-

ulations and has been compared also to several simple (re-)weighting schemes to get an unbiased estimate with less computational costs: their performance is very poor compared to our method.

Nucleation dynamics in 2D cylindrical Ising models and chemotaxis

C.Bosia[1,2], M.Caselle[1,2] and D.Corá[2,3]

[1] Dipartimento di Fisica Teorica dell'Università di Torino and I.N.F.N., via P.Giuria 1, I-10125 Torino, Italy
[2] Center for Complex Systems in Molecular Biology and Medicine, University of Torino, Via Accademia Albertina 13, I-10100 Torino, Italy
[3] Systems Biology Lab, Institute for Cancer Research and Treatment (IRCC), School of Medicine, University of Torino, Str. Prov. 142 Km. 3.95, I-10060 Candiolo, Torino, Italy

Abstract

The aim of our work is to study the effect of geometry variation on nucleation times and to address its role in the context of eukaryotic chemotaxis (i.e. the process which allows cells to identify and follow a gradient of chemical attractant). As a first step in this direction we study the nucleation dynamics of the 2d Ising model defined on a cylindrical lattice whose radius changes as a function of time. Geometry variation is obtained by changing the relative value of the couplings between spins in the compactified (vertical) direction with respect to the horizontal one. This allows us to keep the lattice size unchanged and study in a single simulation the values of the compactification radius which change in time. We show, both with theoretical arguments and numerical simulations, that squeezing the geometry allows the system to speed up nucleation times even in presence of a very small energy gap between the stable and the metastable states. We then address the implications of our analysis for directional chemotaxis. The initial steps of chemotaxis can be modelled as a nucleation process occurring on the cell membrane as a consequence of the external chemical gradient (which plays the role of energy gap between the stable and metastable phases). In nature most of the cells modify their geometry by extending quasi-onedimensional protrusions (filopodia) so as to enhance their sensitivity to chemo-attractant. Our results show that this geometry variation has indeed the effect of greatly decreasing the timescale of the nucleation process even in presence of very small amounts of chemo-attractants.

Correspondence Between Discrete and Continuous Models of Gene Regulatory Network

Aparna Das[1], Francine Diener[1], Gilles Bernot[2], Jean-Paul Comet[2]

[1] Lab. J.A. Dieudonne, Department of Mathematics, University of Nice Sophia-Antipolis, 28, Avenue Valrose, 06108 Nice Cedex-2, France
[2] Lab. I3S, UMR 6070 UNS and CNRS, Algorithmes-Euclide-B, 2000 route des Lucioles, B.P. 121, F-06903, Sophia-Antipolis, France

Abstract

We know that some proteins can regulate the expression of genes in a living organism. The regulation of gene expression occurs through networks of regulatory interactions in a non linear way between DNA, RNA, Proteins and some molecules, called genetic regulatory networks. It is becoming clear that mathematical models and tools are required to analyse these complex systems.

In the course of his study on gene regulatory networks R. Thomas proposed a discrete framework that mimics the qualitative evolution of such systems. Such discrete models are of great importance because kinetic parameters are often non measurable *in vivo* and because available data are often of qualitative nature. Then Snoussi proved consistency between the discrete approach of R. Thomas and Piecewise Linear Differential Equation Systems, which are easy to construct from interaction graph and thresholds of interactions. There exists a transition between two qualitative states (in the discrete model) if and only if there exists a trajectory of the differential model that goes from a point of the domain corresponding to the first qualitative state to the boundary separating this domain to the one corresponding to the second qualitative state.

Our work focuses also on the relationships between both approaches: we would like to extend the result due to Snoussi. Can we give some conditions on the model or on the trace of the qualitative state space which ensures that it is possible to construct a trajectory of the differential model that passes through the same sequence of domains ?

Our main result consists in a theorem stating that, considering a continuous model, for which the associated discrete model has a finite path $s_0 \rightarrow s_1 \rightarrow ... \rightarrow s_n$ such that for all $i \in [1, ..., n-1], s_{i-1} \neq s_{i+1}$, then, under some hypotheses, trajectories of the differential system starting from the domain associated with s_0 pass successively through each domain associated with the states of the path. The used hypotheses have been introduced by Jean-Luc

Gouze and Farcot in a previous work concerning limit cycles. The proof is done by induction and sketch of the proof is given.

Finally, several well chosen examples will illustrate the use of the theorem as well as its limitations due to the stated hypotheses.

CCAS: Cell Cycle Analyser Software

Behzadi M[1], Dupont PY[2], Tran Thuong VD[1], Demidem A[2], Morvan D[2], Schwartz L[1], Souiller H[3], Dantal Y[3], Stepien G[1] and Steyaert JM[1]

[1] LIX, École polytechnique Palaiseau
[2] INRA UMR1019, Clermont-Ferrand
[3] Soluscience SA, Clermont-Ferrand

Abstract

The techniques most frequently used in biology for the study of metabolism (nuclear magnetic resonance (NMR), mass spectrometry (MS), chips, etc) only provide partial information about networks. Then the expert has to reconstruct networks between metabolites for which he has information. In addition, each of these techniques has its limits and only studies a particular aspect of metabolism. By combining the results of these different techniques, we hope to overcome these limits and to derive new and more comprehensive results.

Our goal is to develop a new innovative software to merge data from exploration techniques for the largescale cellular metabolism. This strategy will model the evolution of cellular metabolism as a result of pathology or a particular physiological environment. The data that we have are of four types: Promotology, Transcriptomics, Metabolomics and Proteomics. The intersection of the results of these techniques should help us to overcome some of their limits and to derive the new information. This software will allow to rebuild all possible reactions between the metabolites of interest (identi?ed by NMR spectra or MS, DNA chips, or promotology). The software will be based on the public database KEGG. The database will also record the results of simulation. These results will be then exported in the standard file formats used in modeling : SBML (Biomodels public database) or BioPax (Online database Biological Pathway Exchange).

From the complex metabolic network established by the software, the system of ordinary differential equations which corresponds to this network is automatically created and solved using scilab. The user may choose the biochemical laws which will be used for the construction of this system. One can also enter the parameter values which are known . The unknown parameters are predicted by the software with the numerical methods and will be presented in a table. The results of the simulation of systems and different analyses are presented as diagrams: changes in metabolite concentrations over time,

evolution of concentration of a metabolite in relation to another, phase spaces, stability curves of the system, rate of reaching the steady state for each of the metabolites, etc.

We plan to include other simulation methods or study of biological networks such as calculating elementary flux modes or such as Petri nets.

PreCisIon: PREdiction of CIS-regulatory elements improved by gene's positION

Mohamed Elati[1], Joan Hérisson[1], Rim Fekih[1], Ivan Junier[1], François Képès[1]

[1] iSSB/EPG, Genopole® - CNRS UPS3201 - Univ. Évry Val d'Essonne, F-91030 ÉVRY cedex

Abstract

Transcriptional interactions, occurred between transcription factors (TFs) and their target genes, control many important processes, such as critical steps in development and responses to environmental stresses, and their defects can result in various diseases. To have a deeper understanding of these interactions, the accurate computational prediction of the location of DNA binding sites is therefore a highly desirable research goal, and a key step towards the ability to reverse engineer genetic regulatory networks at a genomic scale. Thus, many algorithms have been developed to exploit the various sources of experimental information available and the various statistical properties that appear to distinguish regulatory regions from the genome in general. However, all these approaches looked for improvement of methods which all rely only on local sequence information, while, to achieve a qualitative jump in the area of binding site prediction would require essentially information of a conceptually novel type. As chromosomal architecture is an essential ingredient for proper transcriptional function [1], it could potentially be used for transcriptional network inference.

Sequence classifier. Recurrent binding sites, in a collection of DNA sequences (promoter region of target's genes), are most commonly modeled by position weight matrices (PWMs). The sequence patterns are simply strings over the 4-letter alphabets $[A, C, G, T]$ that form the DNA. So an L-long sequence motif can be represented by a $4 \times L$ matrix with weights giving the frequency of the four DNA bases in each of the L positions [3]. Discovery of a PWM in sequence data was an early problem to be addressed in computational biology. The challenge is to find the location of the sites and the representative PWM using only the sequence data, without any assumptions on the statistical distributions of patterns in the sequences. Many alignment driven algorithms have been developed.Given a new sequence s and a learned PWM, one can score any subsequence in that input.

Positional classifier. Recent post-genomics studies have unrevealed regular patterns in the position of some genes along the DNA [1]. Two types of patterns have been identified for genes that are regulated by a common TF: they tend either to be clustered along the DNA, hereafter referred to as 1D clustering, or to be placed all along the genomes according to a periodic organization. As a consequence, the position of the genes shall be seeing as an

expedient classifier for the prediction of TF binding sites. Here, we propose to use both the 1D clustering information and the periodic trends to generate positional score as an additional classifier. Within this scope, authors of this paper have recently developed a tool for analyzing the periodic trends in small sets of positional data points (discrete one-dimensional signal). The tool aims i) to detect the presence of a periodic pattern and ii) to allocate to each point (gene) a positional score, which reflects the tendency of the point to participate to the periodic trend.

Classifier fusion using boosting. In principle there are two approaches to combining classifiers, namely classifier fusion and classifier selection. But, only classifier fusion methods were explored in our model seen that he individual classifiers described above were designed to be global experts. In this paper we propose a slight variation of the AdaBoost algorithm [2] for our classifier fusion problem. AdaBoost has been shown to improve the prediction accuracy of weak classifiers using an iterative weight update process. The technique combines weak classifiers (classifiers having classification accuracy slightly greater than chance) in a weighted vote fashion giving an overall strong classifier. Our variation to the regular form of AdaBoost consists in allowing the algorithm to choose, in each iteration, among weak classifiers trained on different views (sequence and position in our case) of the training data. The combination weights for the final weighting rule are obtained using a shared sampling distribution. In each iteration, a weak classifier is greedily selected from the pool of weak learners trained on disjoint views. This results in a minimization of the training error for the final hypothesis. Notice that while this is not the regular procedure for training AdaBoost, we are not modifying any of the assumptions that the algorithm is based on; we only extend its hypothesis space. As a result, all the convergence proofs that apply to AdaBoost also apply to our version of the algorithm.

Results. In order to assess the performance of the proposed method, and compare it with other existing methods, we test it on a various transcription factors of *E. coli* (from RegulonDB) and *B. Subtilis* (from DBTBS) using a cross-validation procedure. The results show that our method tends to improve the prediction of target and non-target genes compared to the individual classifiers.

References

[1] Képès F. (2004). Periodic transcriptional organization of the e.coli genome. *J Mol Biol*, **340**, 957–964.

[2] Schapire R. E. (1999) A brief introduction to boosting. In *IJCAI*, p. 1401–1406.

[3] Stormo G. D. & Hartzell G. W. (1989). Identifying protein-binding sites from unaligned dna fragments. *PANS*, **86**(4), 1183–1187.

Signatures of chemical diversity in metabolic networks

Davide Fichera[1], Pablo Carbonell[1], Jean-Loup Faulon[1]

[1] iSSB, Institute of Systems and Synthetic Biology, University of Evry, Genopole, 91030 Evry cedex, France

Abstract

The plasticity mechanisms that make catalytic active sites able to process different biochemical reactions, and to accommodate different substrates, are at present not fully understood [1]. In this work, we will present recent advances in the use of a graph-based representation known as molecular signature [2]. Molecular signatures provide a description of the atoms and their environment in the molecule. This characterization fully identifies chemical species in order to explore the chemical space at both sides of the enzyme-substrate binding process [3]. Thus, this technique characterizes both chemical species and their reactions within a common framework, allowing us to study the relationship between properties of chemical structures such as stereochemistry or molecular similarity, and the process of substrate recognition, catalytic specificity efficiency of the reaction, as well as biological activity.

Performing this study for the entire KEGG metabolic database [4] brings insights into evolutionary relationships between different metabolic pathways, and how new catalytic functions are acquired. We show here enzyme engineering applications of this method to the directed evolution of natural occurring protein scaffolds with latent catalytic activities.

References

[1] Nobeli, I., Favia, A. D., and Thornton, J. M. (2009). Protein promiscuity and its implications for biotechnology. *Nature Biotechnology*, **27**(2), 157–167.

[2] Faulon, J.-L. L., Collins, M. J., and Carr, R. D. (2004). The signature molecular descriptor. 4. canonizing molecules using extended valence sequences. *J. Chem. Infor. Comp. Sci.*, **44**(2), 427–436.

[3] Faulon, J.-L., Misra, M., Martin, S., Sale, K., and Sapra, R. (2008). Genome scale enzyme metabolite and drug target interaction predictions using the signature molecular descriptor. *Bioinformatics*, **24**(2), 225–233.

[4] Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., and Yamanishi, Y. (2008). Kegg for linking genomes to life and the environment. *Nucleic acids research*, **36**(Database issue), D480–484.

MetaboFlux : a method to analyse flux distributions in metabolic networks

Amine Ghozlane[1,2], Frédéric Bringaud[3], Fabien Jourdan[4] and Patricia Thébault[1,2]

[1] CBiB, Univ. Bordeaux 2
[2] LaBRI, Univ. Bordeaux 1
[3] RMSB, UMR 5536 CNRS, Univ. Bordeaux 2
[4] INRA, UMR 1089 Xénobiotiques, Toulouse, France

Abstract

*T*rypanosoma brucei is a parasitic protist of vertebrates that causes sleeping sickness in Africa. A part of its energetic metabolism, including 7 of the early stages of glycolysis, occurs in an organelle called glycosome. A metabolic pathway for the glycosome had been built by exploiting genomic, proteomic and metabolomic data [1].

Some known biological constraints, such as the ATP/ADP and NADH/NAD+ balances, are not considered in the model. We propose a modelling approach including structural pathway and metabolic flux analysis to help in the understanding of the system's structure and its semi-quantitative behaviour.

We model known biological information with a stochastic Petri net (where transitions are given for the reaction and places for metabolites) where delays can be assigned to transitions given a probability distribution. From a given set of probability distribution representing the flux amount of reactions (the input set of parameters), the simulation of the Petri net allows the exploration of the possible behaviours of the system. At the end of a run, if all input metabolites are consumed, we get concentration for intermediate and output metabolites. We integrate expected metabolites concentrations revealed by biological experiments within an objective function, and use simulated annealing and simplex minimization approach for its global optimization. Therefore, simulations are carried out by fitting the set of input parameters until the system reach the best optimization of the objective function. To explore a large set of possible behaviour of the system, several run of simulations combined with the simulated annealing approach are made. A set of solutions is given by different groups of fluxes distributions (that best fit expected metabolites concentrations), and are helpful to make some assumptions and analysis for a given metabolic system.

MetaboFlux was developed to this purpose and applied to *T*. brucei. Resulting scenarios strongly argue in favour of an unrealistic NADH/NAD+ imbalance and suggest adding to the model new metabolic pathways. A realistic solution may be to integrate the pentose phosphates to the previous model. The resulting new model was tested with Metaboflux and shows relevant fluxes scenarios.

References

[1] Coustou V., Brian M.B., Breton M., Guegan F., Rivière L., Plazolles N., Nolan D., Barrett M.P., Franconi J.-M., Bringaud F., (2008) Glucose-induced remodelling of intermediary and energy metabolism in procyclic T. brucei. *J.B.C.* **283**: 16342-16354.

Investigation of the topological structure of global genetic networks

Svilen Iskrov[1], Jean-Baptiste Rouquier[2], Laurent Jannière[3], François Képès[3] and Ivan Junier[2,3]

[1] M2 Physique des Liquides UPMC/ ENS ULM/ Polytechnique, Paris, Ile-de-France
[2] UMR7656 CREA CNRS, Institut des Systemes Complexes, Paris, Ile-de-France
[3] Programme d'Épigénomique - Genopole, CNRS UPS3201, PRES UniverSud Paris, UEVE Genopole Campus 1 - Genavenir 6, 5 rue Henri Desbruères F-91030 EVRY cedex

Abstract

There is increasing evidence that the overall regulation of genetic expression relies on a proper spatio-temporal organization of chromosomes inside the cells. This can be seen in particular from the transcription machinery: in some eukaryotes and bacteria, transcription of highly active (co-functional or co-regulated) genes occurs within discrete foci called transcription factories, where RNA polymerases, transcription factors and their target genes co-localize [1] . In my work, I will show that DNA inter-gene distances can be used to efficiently work out functional links between genes. In particular, I will present a network – the closome – that reflects the tendency for COGs (cluster of orthologous genes) to be close to each other along the DNA. It has been obtained by analyzing the genomic organization of more than 800 different bacteria. Various properties like degree distribution, betweenness and community structure [2] are revealed and discussed.

References

[1] Junier, I., Martin, O., Képès, F. (2010), Spatial and topological organization of DNA chains induced by gene co-localization, *PLoS Comput. Biol.* **6(2)**: e1000678 .

[2] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E. (2008), Fast Unfolding of Community Hierarchies in Large Networks. *Eprint arXiv:0803.0476v1* at arxiv.org

Coupling between metabolism and replication depends on the flux travelling terminal reactions of glycolysis

Hamid Nouri[1], Anne-Françoise Monnier[1], Monika Maciag[2], Solveig Fossum[3], Armelle Cabin-Flaman4, Kirsten Skarstad[3], Gregorz Wegrzyn[2], Agnieszka Szalewska-Palasz[2] and Laurent Jannière[1]

[1] Laboratoire de Génétique Microbienne, INRA, 78350 Jouy en Josas, France
[2] University of Gdansk, Department of Molecular Biology, Kladki 24, 80-822 Gdansk, Poland
[3] Department of Cell Biology, Institute of Cancer Research, Norwegian Radium Hospital, Rikshospitalet, University of Oslo, 0310 Oslo, Norway
[4] Laboratoire d'Assemblages moléculaires: modélisation et imagerie SIMS, Faculté des Sciences de l'Université de Rouen, F-76821 Mont Saint Aignan Cedex, France & DYCOEC, CNRS (GDR 2984), CORIA, Université de Rouen, France

Abstract

DNA synthesis is coupled to the growth rate afforded by nutrients by an unknown mechanism. To evaluate the involvement of metabolic reactions in this process, we investigated the relationship between growth and replication in *Bacillus subtilis* metabolic mutants. For this, the ratio of origin to terminus chromosomal sequences was plotted against the number of cell doublings per hour. In $\Delta ptsG$, Δpgi, $pfk-1$, $\Delta citC$, Δmdh, $\Delta pdhB$ and $\Delta pdhC$, a WT relationship was observed and an increase in origin sequences occurred in $\Delta gapA$, $pgk-EP$ and $\Delta pykA$ strains. Flow cytometry analysis showed that the increase is due to a defect in replication control. A similar phenotype was observed in cells deleted for genes playing moderate ($pycA$) or no known metabolic function ($gapB$, $pckA$) in LB and in cells deleted for the regulators of polarity and intensity of the flux travelling terminal reactions of glycolysis ($ccpN$ and $yqfL$). However, no replication phenotype was observed in mutants of global metabolic transcriptional regulators ($ccpA$ and $codY$). Results also showed that some combination of carbon sources can deregulate replication in $\Delta pykA$ and $\Delta pckA$ mutants.

We conclude that (i) coupling is driven by the carbon flux passing through terminal reactions of glycolysis, (ii) side reactions connecting the bottom part of glycolysis to surrounding paths ($pycA$, $gapB$ and $pckA$ in glycolytic nutrients and $pykA$ in neoglucogenic regimen) balance glycolytic flux by carrying out anaplerotic or cataplerotic reactions and that (iii) unbalanced flux alters coupling and deregulates replication. These results support our previous observations (*PLoS ONE* (2007) **2(5)**: e447).

Molecular networks and modeling with the Biocham Graphical User Interface

Dragana Jovanovska[1]

[1] Biocham Project, CONTRAINTES Team, INRIA Paris Rocquencourt, France

Abstract

The biochemical abstract machine BIOCHAM is a modelling environment based on two languages for formalizing respectively, the elementary interactions between compounds, and the system behaviour under different conditions. The Biocham GUI is its graphical user interface that gives richer user experience, making it more easy, effective and animated. It implements all the functional features that the modelling environment Biocham offers and more others, including: the comparison section for results from multiple different numerical simulations or traces; customizing results preview features, automatic creation of LTL queries by giving data file input, etc... Biologists use diagrams to represent interactions between molecular species, and on the computer, diagrammatic notations are also more and more employed in interactive maps. Therefore, the Biocham GUI is being enriched with a graphical reactions editor for constructing and editing biochemical reactions, and systems of biochemical reactions, with kinetic expressions, as written in the Systems Biology Markup Language SBML, and interpreted by a system of Ordinary Differential Equations over molecular concentrations. The Biocham reactions graphical editor supports the SBGN specification for a graphical notation. Today, we are discovering different graph layouts that will make the complex reactions graphs readable for the users and that will enable the users find the necessary information from the graphical representation of the molecular interactions more quickly and precisely.

References

[1] François Fages and Sylvain Soliman, Formal Cell Biology in BIOCHAM (tutorial) 8th International School on Formal Methods for the Design of Computer, Communication and Software Systems: Computational Systems Biology. In memory of Nadia Busi. Bertinoro, Italy. Springer-Verlag, LNCS 5016. Mar. 2008

[2] François Fages and Sylvain Soliman, From reaction models to influence graphs and back: a theorem Formal Methods in Systems Biology. Springer-Verlag LNCS. Feb. 2008

Using Hoare logic for constraining parameters of discrete models of gene networks

Zohra Khalis[1,2], Gilles Bernot[1], Jean-Paul Comet[1]

[1] Lab. I3S, UMR 6070 UNS and CNRS, Algorithmes-Euclide-B,
2000 route des Lucioles, B.P. 121, F-06903, Sophia-Antipolis, France
[2] Programme d'Épigénomique - Genopole, CNRS UPS3201, PRES UniverSud Paris,
UEVE Genopole Campus 1 - Genavenir 6, 5 rue Henri Desbruères
F-91030 EVRY cedex

Abstract

The technology of DNA chips allows the quantification of the expression level of thousands genes in the same organism at the same time. Nevertheless, analysis of data from DNA chips requires development of adapted methods.

We propose a path language that allows the description, in an abstract way, of the concentration level variations from temporal data like temporal profiles of gene expression. When concentration level variations have been expressed through a program of the path language, it becomes possible to apply some methods from computer science like Hoare logic.

Hoare logic is made of a system composed of axioms and rules. It permits one to prove if a program is correct in comparison to its specification that is described through assertions, that is, logical formulas, that define properties on the program. The precondition specifies the initial state before the execution of the program and the postcondition specifies the final state after the execution of the program. A program is said (partially) correct if it is possible to prove that from an initial state satisfying the precondition, the program leads (if the program terminates) to a final state satisfying the postcondition.

To model gene regulatory networks, the main difficulty remains in the parameter identification problem, that is, the search of parameter values which lead to a model behavior that is consistent with the known or hypothetical properties of the system. So, we apply a Hoare-like logic designed for the defined path language. The axioms and rules of this Hoare-like logic are adapted to gene networks and permits us to prove that the path described by the program exists in the dynamics. Given a path program and a postcondition, we can apply calculus of the weakest precondition, based on this Hoare-like logic. Calculus of the weakest precondition, thanks to defined axioms and rules, permits us to constrain parameters associated with the program and the postcondition. Although Hoare logic is well known, its application to constrain

parameter values of gene networks appears to be brand new and helpful in order to select consistent models. Moreover, expressing DNA profiles into programs gives another way to process such data.

Bipartite Graph Properties and Systems Biology

Faten Nabli[1], Sylvain Soliman[1] and François Fages[1]

[1] Project team CONTRAINTES, INRIA Paris-Rocquencourt, France

Abstract

Since 1993, Petri Nets have emerged as a promising tool to describe and analyze biochemical networks combining the qualitative and the quantitative approaches.

In our work, we explore Petri Nets structural properties to conclude about the biochemical system evolution and its dynamics especially in steady state conditions.

Our motivation to study structural properties is due to the fact that they can represent the unique way to extract some information about the dynamics of the biochemical system because of the lack of the kinetic data necessary to establish ODEs of certain big biochemical networks.

T-invariant computation is already frequently used in flux distribution analysis since minimal T-invariant correspond to elementary modes that are defined as the minimal set of reactions that can occur at steady state.

In our case, in addition to flux distribution, we are also interested in a less abstracted view of models and we search a way to compute the chemical components concentrations at steady state. In this field we achieve some results concerning the computation of some steady states based on T-invariant search of the underlying Petri Net.

Modelling TGFβ-dependent NFκB response in cervical cancer cell-lines

Vincent Noël[1], Ovidiu Radulescu[2], Tessy Maliekal[3]

[1] RMAR, Université de Rennes 1
[2] DIMNP, Université de Montpellier 2
[3] NCBS, Bangalore

Abstract

NFκB signalization pathway is a key component of our immune system, that can be activated by various stimuli such as bacterial or viral products, or stress. Activation of this pathway can induce expression of more than 500 genes in a temporally controlled manner. This pathway is an important intermediate, and his regulation remains unclear despite years of research.

We model the dose-dependent induction of epithelial-mesenchymal transition (EMT) in response to a TGFβ treatment. An experiment with NFκB overexpressing cells proves that NFκB plays a key role in this dose-dependent response. The purpose of this study is to test mechanistic hypothesis about how TGFβ acts on NFκB pathway, and to propose experiments to help biologists investigating this crosstalk.

Using the mathematical framework we showed that several biochemical mechanisms involving simple and multiple acetylation of NFκB allow to explain the dose responding behavior and we propose the following experimental tests to discriminate between mechanistic hypothesis:

- Measuring acetylation status of p65 on lysine 211 and 310 and its influence on IκBα - NFκB binding reaction
- Probing IκBα, A20 and IκBα kinase activity

These results are important for understanding pathway cross-talk in cancer. Future experimental and theoretical work should elucidate how different NFκB nuclear signals can trigger different gene responses. A possible simple hypothesis is based on very different time scales of genes. Slower genes respond to the average NFκB level which is lower in the oscillation with respect to the non-oscillating signal, while rapid genes respond to the transient peak of activity.

The role of incoherent microRNA-mediated feedforward loops in noise buffering.

Matteo Osella[1,2], Carla Bosia[1,2], Davide Cora'[1,2] and Michele Caselle[1,2]

[1] Dipartimento di Fisica Teorica and INFN University of Torino, v. Pietro Giuria 1, 10125 Torino, Italy.
[2] Center for Complex Systems in Molecular Biology and Medicine University of Torino, Via Accademia Albertina 13, I-10100 Torino, Italy.

Abstract

MicroRNAs are endogenous non coding RNAs that play important gene regulatory roles in animals and plants by pairing to the messenger RNAs of protein-coding genes to direct their post-transcriptional repression. Transcriptional and miRNA regulations are interlinked in a complex network in which the microRNA-mediated feed forward loop seems to be a motif (an overrepresented regulatory circuit). We show analytically and through simulations that the incoherent version of this circuit can couple a fine-tuning of the target protein level with a noise buffering. In particular it can confer robustness to the gene expression of the target with respect to fluctuations in upstream factors. Moreover our model predicts that the optimal attenuation of fluctuations coincides with a modest repression of the target expression. This feature is coherent with a fine-tuning function and in agreement with experimental observations of the actual impact of a wide class of microRNAs on the protein output of their targets.

BioΨ: a formal description of biological processes based on elementary bricks of actions

Sabine Pérès [1], Liza Felicori [1], Stéphanie Rialle [1] and Franck Molina[1]

[1] Sysdiag CNRS Bio-Rad UMR 3145, Cap delta/Parc euromédecine,
1682 rue de la Valsière CS 61003, 34184 Montpellier Cedex 4 France.

Abstract

In the available databases, biological processes are described from molecular and cellular points of view, but these descriptions are represented with text annotations that make difficult to handle them for computation. Consequently, there is an obvious need for formal descriptions of biological processes. A formalism that uses the BioΨ concepts to model biological processes from molecular details to networks will be presented. This computational approach, based on elementary bricks of actions, allows us to calculate on biological functions (e.g. process comparison, mapping structure-function relationships, etc.). Its application will be illustrated with the functional description of the central carbon metabolism network. This computational approach is compatible with detailed biological knowledge and can be applied to different kinds of systems of simulation.

Refining Dynamics of Gene Regulatory Networks in a Stochastic π-Calculus Framework

Loïc Paulevé[1], Morgan Magnin[1], Olivier Roux[1]

[1] IRCCyN, UMR CNRS 6597, École Centrale de Nantes, France
{pauleve,magnin,roux}@irccyn.ec-nantes.fr

Abstract

We introduce a framework, the Process Hitting [1], allowing to model and analyse efficiently Gene Regulatory Networks (GRNs) in their temporal and stochastic aspects.

Starting from a GRN without any other parameters, its largest dynamics (or *generalised dynamics*) are expressed in Process Hitting and then are refined to match the expected behaviour. Such a refinement is achieved by constructing cooperativity between genes and by creating stable states. The analysis of stable states and inference of René Thomas' discrete parameters derives from this logical formalism.

This framework comes with a natural translation to the Stochastic π-Calculus bringing time and stochasticity into the models. Efficient simulations through SPiM [2] and probabilistic model checking using PRISM [3] are then possible. The merits and scalability of our framework is illustrated on the control of the differentiation in a GRN generalising metazoan segmentation processes [4], and on the analysis of stable states within a large GRN studied in the scope of breast cancer researches [5].

References

[1] Paulevé, L., Magnin, M., Roux, O.: Refining Dynamics of Gene Regulatory Networks in a Stochastic π-Calculus Framework. Research Report hal-00397235 (June 2009)

[2] Phillips, A., Cardelli, L.: Efficient, correct simulation of biological processes in the stochastic pi-calculus. In: Computational Methods in Systems Biology. Volume 4695 of LNCS., Springer (2007) 184–199

[3] Hinton, A., Kwiatkowska, M., Norman, G., Parker, D.: PRISM: A tool for automatic verification of probabilistic systems. 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06). Volume 3920 of LNCS., Springer 441–444

[4] Francois, P., Hakim, V., Siggia, E.D.: Deriving structure from evolution: metazoan segmentation. Mol Syst Biol **3** (Dec 2007)

[5] Sahin, O., Frohlich, H., Lobke, C., Korf, U., Burmester, S., Majety, M., Mattern, J., Schupp, I., Chaouiya, C., Thieffry, D., Poustka, A., Wiemann, S., Beissbarth, T., Arlt, D.: Modeling erbb receptor-regulated g1/s transition to find novel targets for de novo trastuzumab resistance. BMC Systems Biology **3**(1) (2009)

GraMoFoNe: a Cytoscape plugin for querying motifs without topology in PPI networks

Guillaume Blin, Florian Sikora and Stéphane Vialette

Université Paris-Est, LIGM - UMR CNRS 8049, France
{gblin, sikora, vialette}@univ-mlv.fr

Abstract

During the last decade, data on Protein-Protein Interactions (PPI) has increased in a huge manner. Searching for motifs in PPI Network has thus became a crucial problem to interpret this data. A large part of the literature is devoted to the query of motifs with a given topology. However, in some situations, the topology is not known or is irrelevant, which leads to searching *functional* motifs instead of *topological* ones.

In this setting, we still ask for the conservation of the node labels, but we replace topology conservation by the weaker requirement that the subnetwork should form a connected subgraph of the target graph. This approach was advocated by [2] and led to the definition of the GRAPH MOTIF problem: given a vertex-colored graph $G = (V, E)$ and a multiset of colors M, find a set $V' \subseteq V$ such that the induced subgraph $G[V']$ is connected, and the multiset of colors of the vertices of V' is equal to M. In our context, the graph G represents the PPI network where vertices are the proteins and edges the interactions. The motif is completely defined by adding a color in M for each different requested proteins. A node $v \in G$ is colored by a color $c \in M$ if the protein represented by v is homologous to the protein represented by c

Despite the NP-completeness of the problem [2], a lot of theoretical results exists. In this contribution, we present GraMoFoNe, a plugin to Cytoscape based on a Linear Pseudo-Boolean optimization program (*i.e.*, as a linear program whose variables are boolean) which handles GRAPH MOTIF and some of its extensions (to be published in [1]).

Moreover, for a large scale test purpose, we also provide a batch mode of our plugin. We used this last to retrieve motifs (protein complexes) of six different species in three large different PPI networks. Comparisons between our plugin GraMoFoNe and Torque were computed. For most experiments, our plugin finds more feasible motifs and also more "true solutions" than Torque. The GraMoFoNe plugin are available at
`http://igm.univ-mlv.fr/AlgoB/gramofone` or through Cytoscape plugin page.

References

[1] G. Blin, F. Sikora, and S. Vialette. GraMoFoNe. In *BICoB*. ISCA, 2010.

[2] V. Lacroix, C. Fernandes, and M.-F. Sagot. Motif search in graphs: application to metabolic networks. *TCBB*, **3(4)**:360-368, 2006.

Using Multi-Agent Systems to Design Internal Movements of Proteins

François Vallée[1], Marie Beurton-Aimar[1], Nicolas Parisey[2]

[1] LaBRI, Unité Mixte de Recherche CNRS (UMR 5800), Université Bordeaux 1, Université Bordeaux 2

[2] UMR Bio3P INRA - le Rheu

Abstract

Multi-Agent Systems (MAS) are systems composed of multiple autonomous entities (agents) which can collaborate. Applied to biological issues, they were used to model ecological systems such as insect colonies where agents can easily model the simple relationships between entities and the complex behaviour of the entire system. In the field of cellular and molecular biology, an agent is often a cell but rarely a single molecule. Moreover, these kinds of agents are not influenced by their shape or their physical features as they generally are non-physical agents. To model biological processes at the molecular level, especially enzymatic reactions, we have to take into account physical properties and explicit three-dimensional representation of molecules. Thus, we have defined a type of reactive agents composed of rules which correspond to the behaviour of agents to model enzymatic reactions and of a body which consists in the three-dimensionnal representation and the rules for physical interactions and internal movements. In our model, we also have two subtypes of reactive agents: active agents and passive agents which are respectively enzymes and substrates because they do not have common rules regarding to enzymatic reactions and their spatial representations are extremely different. For enzymes, three-dimensionnal structures are given by structural data extracted from the Protein DataBank (PDB). To determine possible internal movements, we have analyzed these data with hinge determination algorithms and elastic network models to identify rigid parts of macromolecules and thus the possible conformational changes. The first application to respiratory chain complexes allows us to split these macromolecules in few parts (3 or 4) and we obtained a simple yet realistic representation. From this model, we are able to create simulations of electron chain transport which include both information, those from the movements of the enzyme body and those from the biochemical reaction. Moreover the influence of the movements on the reaction efficiency is directly available in this kind of simulations. In future works, models will have to take into account interactions between molecules which have not the same size like enzymatic complexes and phospholipids. To work, our model would be able to adapt the physical interactions to multi-scale modelling.

LIST OF ATTENDEES

(March 21st, 2010)

ABCHICHE-MIMOUNI	Nadia (nadia.abchiche@ibisc.univ-evry.fr)
AMAR	Patrick (pa@lri.fr)
AKUTSU	Tatsuya (takutsu@kuicr)
ARACENA	Julio (jaracena@ing-mat.udec.cl)
BEN JACOB	Eshel (eshel@tamar.tau.ac.il)
BERNOT	Gilles (bernot@unice.fr)
BERTOLINO	Arianna (arianna.bertolino@gmail.com)
BEURTON-AIMAR	Marie (beurton@labri.fr)
BONNEUIL	Christophe (bonneuil@damesme.cnrs.fr)
BOSIA	Carla (cbosia@to.infn.it)
CHANDESRIS	Jérôme (jerome.chandesris@unice.fr)
COMET	Jean-Paul (comet@unice.fr)
DAS	Aparna (aparnadas143@yahoo.com)
DUPONT	Pierre-Yves (pierre-yves.dupont@clermont.inra.fr)
ELATI	Mohamed (mohamed.elati@epigenomique.genopole.fr)
FANCHON	Éric (eric.fanchon@imag.fr)
FICHERA	Davide (fichera.davide@gmail.com)
FOX	Gavin (fox@ibs.fr)
FROMENTIN	Jonathan (jonathan.fromentin@u-bordeaux2.fr)
FROMION	Vincent (Vincent.Fromion@jouy.inra.fr)
GAZÈRES	Nicolas (ngs@3ds.com)
GHOZLANE	Amine (amine.ghozlane@labri.fr)
GLATIGNY	Annie (annie.glatigny@orange.fr)
GUESPIN	Janine (janine.guespin1@orange.fr)
HAMID	Nouri (hamid.nouri@epigenomique.genopole.fr)
HATZIMANIKATIS	Vassily (vassily.hatzimanikatis@epfl.ch)
van HELDEN	Jacques (jacques.van.helden@ulb.ac.be)
HÉRISSON	Joan (Joan.Herisson@epigenomique.genopole.fr)
ISKROV	Svilen (svilen.iskrov@gmail.com)
JACQUET	Michel (michel.jacquet@igmors.u-psud.fr)
JANNIÈRE	Laurent (laurent.janniere@epigenomique.genopole.fr)
JARAMILLO	Alfonso (Alfonso.Jaramillo@epigenomique.genopole.fr)
JOVANOVSKA	Dragana (dragana.jovanovska@inria.fr)

JUNIER	Ivan (i.junier@gmail.com)
KAETZEL	Thorben (thorben.kaetzel@uni.lu)
KÉPÈS	François (francois.kepes@epigenomique.genopole.fr)
KHALIS	Zohra (khalis@i3s.unice.fr)
KLECKNER	Nancy (kleckner@fas.harvard.edu)
LARIVIÈRE	Damien (damien.lariviere@fourmentinguilbert.org)
LEMAIRE	Sophie (sophie.lemaire@math.u-psud.fr)
LOPEZ	Alvaro (alvarojh123@gmail.com)
MAZAT	Jean-Pierre (Jean-Pierre.Mazat@phys-mito.u-bordeaux2.fr)
MOLINA	Franck (franck.molina@sysdiag.cnrs.fr)
MUELLER	Melanie (mmueller@mpikg.mpg.de)
NABLI	Faten (faten.nabli@inria.fr)
NÉDÉLEC	François (nedelec@embl.de)
van NIMWEGEN	Erik (erik.vannimwegen@unibas.ch)
NOËL	Vincent (labestiol@gmail.com)
NORRIS	Victor (victor.norris@univ-rouen.fr)
OSELLA	Matteo (mosella@to.infn.it)
PAULEVÉ	Loïc (pauleve@irccyn.ec-nantes.fr)
PERES	Sabine (sabine.peres@sysdiag.cnrs.fr)
RADULESCU	Ovidiu (ovidiu.radulescu@univ-montp2.fr)
RIVAS	Germán (grivas@cib.csic.es)
RIZK	Aurelien (aurelien.rizk@inria.fr)
SACKMANN	Erich (sackmann@ph.tum.de)
SIKORA	Florian (sikora@univ-mlv.fr)
STÜLKE	Jörg (jstuelk@gwdg.de)
THIMO	Rohlf (rohlf@epigenomique.genopole.fr)
TRACQUI	Philippe (philippe.tracqui@imag.fr)
VALLÉE	François (vallee@labri.fr)
WEISSENBACH	Jean (jsbach@genoscope.cns.fr)
WALKER	Michael (michael.walker@sophia.inria.fr)
WONG	Hua (wong.hua@gmail.com)
ZELISZEWSKI	Dominique (dominique.zeliszewski@epigenomique.genopole.fr)
ZEMIRLINE	Abdallah (zemirlin@univ-brest.fr)