Proceedings of the Évry Spring School
on

# advances in Systems and Synthetic Biology

March 21$^{st}$ - 25$^{th}$, 2016

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

Systems Biology includes the study of interaction networks and, in particular, their dynamic and spatiotemporal aspects. It typically requires the import of concepts from across the disciplines and crosstalk between theory, benchwork, modelling and simulation. The quintessence of Systems Biology is the discovery of the design principles of Life. The logical next step is to apply these principles to synthesize biological systems. This engineering of biology is the ultimate goal of Synthetic Biology: the rational conception and construction of complex systems based on, or inspired by, biology, and endowed with functions that may be absent in Nature.

This annual School started in 2002. It was the first School dedicated to Systems Biology in France, and perhaps in Europe. Since 2005, Synthetic Biology has played an increasingly important role in the School. Generally, the topics covered by the School have changed from year to year to accompany and sometimes precede a rapidly evolving scientific landscape. Its title has evolved in 2004 and again in 2012 to reflect these changes. The first School was held near Grenoble after which the School has been held in various locations. It started under the auspices of Genopole®, and has been supported by the CNRS since 2003, as well as by several other sponsors over the years.

This book gathers overviews of the talks, original articles contributed by speakers and students, 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, Attila Csikasz-Nagy, Oliver Ebenhoeh, Ivan Junier,*

*Marcelline Kaufman, François Képès, Pascale Le Gall, Sheref Mansy, Jean-Pierre Mazat, Victor Norris,*

*William Saurin, El Houssine Snoussi, Ines Thiele, Birgit Wiltschi.*

# ACKNOWLEDGEMENTS

# INVITED SPEAKERS

**MADALENA CHAVES**  Biocore group, INRIA Sophia Antipolis, FR

**PAUL FRANÇOIS**  Dept. Physics, McGill U., CA

**EROL GELENBE**  ISN, Imperial College London, UK

**JOHN GLASS**  Synthetic Biology & Bioenergy Group, JCVI, Rockville, US

**JONATHAN KARR**  Inst. for Genomics & Multiscale Biology Inst. Mount Sinaï School of Medicine, New York, US

**TAMAS KORCSMAROS**  GAC / Inst. Food Research, Norwich, UK

**YANNICK RONDELEZ**  LIMMS/CNRS-IIS, U. of Tokyo, JP

**SILVIA SANTOS**  MRC clinical sciences centre, Imperial College London, UK

**STEFAN SCHILLER**  FRIAS, Freiburg, DE

**MARKUS SCHMIDT**  Biofaction, Vienna, AT

**FRIEDRICH SIMMEL**  Systems Biophysics and Bionanotechnology, Technische Universität München, DE

# CONTENTS

## PART II ARTICLES 37

## PART III LIST OF ATTENDEES 142

# Toward whole-cell models for science and engineering

Jonathan KARR[1]

[1] Inst. for Genomics & Multiscale Biology Inst., the Mount Sinai
School of Medicine, New York, US

## *Abstract*

A central challenge in biology is to understand how phenotype arises from genotype. Despite decades of research, a complete understanding of biology remains elusive. Computational techniques are needed to create a unified understanding. Recently, we developed the first comprehensive whole-cell computational model which accounts for the specific function of every annotated gene product and predicts the dynamics of every molecular species over the entire cell cycle. We validated our model by comparing its predictions to a wide range of experimental data. We have used the model to gain new insights into cell cycle regulation and energy usage. We believe that whole-cell models will accelerate bioengineering and medicine by enabling rapid, low cost in silico experimentation, facilitating experimental design and interpretation, and ultimately guiding rational biological design.

# Signaling and regulatory networks

Tamas KORCSMAROS[1]

[1] The Genome Analysis Centre (TGAC) and Institute of Food Research,
Norwich, UK

## *Abstract*

Signaling pathways control most stress response mechanisms. We previously created SignaLink, a signaling resource (`http://signalink.org`) containing manually curated data of major signaling pathways. To provide a detailed mapping of signaling systems we further developed a novel concept to integrate and utilize regulatory mechanisms of the signaling network. The multi-layered (onion-like) database structure of SignaLink 2 is made up of signaling pathways, their pathway regulators (eg, scaffolds) and modifier enzymes (eg, phosphatases, ubiquitin ligases), as well as transcriptional and post-transcriptional regulators (transcription factors and miRNAs). The user-friendly website allows the interactive exploration of how each protein is regulated. We used SignaLink and the multi-layered network concept to investigate the systems-level properties of two stress-related mechanisms: 1) NRF2, a master transcriptional regulator of oxidative and xenobiotic stress responses; 2) autophagy (cellular self-eating) that involves the sequestration and degradation of cytosolic materials.

Autophagy is a complex cellular process having multiple roles, depending on tissue, physiological or pathological conditions. Major post-translational regulators of autophagy are well known, however, they have not yet been collected comprehensively. The precise and context dependent regulation of autophagy necessitates additional regulators, including transcriptional and post-transcriptional components that are listed in various datasets. Prompted by the lack of systems-level autophagy-related information, we developed an online resource, Autophagy Regulatory Network (ARN; `http://autophagy-regulation.org`; Turei at al, Autophagy, 2015), to provide an integrated database for autophagy research. ARN contains manually curated, imported and predicted interactions of autophagy components in humans. We listed transcription factors and miRNAs that could regulate autophagy components or their protein regulators. The user-friendly website of ARN allows researchers without computational background to search, browse and download the database. The database can be downloaded in various file formats. ARN has the potential to facilitate the experimental validation of novel autophagy components and regulators. In addition, ARN helps the inves-

tigation of transcription factors, miRNAs and signaling pathways implicated in the control of the autophagic pathway.

Autophagy is also known to be important for intestinal homeostasis and its malfunction is related to inflammatory bowel disease (IBD). Autophagy is often manipulated by intestinal pathogenic bacteria, such as *Salmonella*. To investigate how *Salmonella* is modulating autophagy we developed the first large-scale network resource for *Salmonella enterica*, integrating known and predicted regulatory, metabolic and signalling interactions. Then, we integrated earlier identified *Salmonella*-host interactions and data from ARN to list and predict novel genes responsible for autophagy modulation in the gut. The developed bioinformatics workflows and experimental validation system could be used for other pathogens.

# Towards imagining Bio-Futures

Markus SCHMIDT[1]

[1] Biofaction KG, Vienna, AT

## *Abstract*

Science and technology and hence also synthetic biology, is informed and propelled by our ability to imagine futures that do not (yet) exist. Without imagination we would not be able to discover new conceptual territories, design innovative products, facilitate new forms of human interaction, etc.

In the third millennium the world is confronted with a number of complex problems and challenges such as increasing demand for energy, water and other natural resources while maintaining an ecologically sustainable economic growth and mitigating global environmental change.

Not unlike the case of the Cartesian split between body and mind that manifests itself in the academic division between science and the humanities, the collaboration between science and various genres of imagination such as fine and visual arts, design, gamification, and (multi-media) storytelling is up for improvement in order to unlock the hidden potential that is required to enable a truly flourishing society.

In this talk I will shed a light on the relationship between science and genres of the imagination such as visualisation of speculative futures, art-science residencies and workshops, reflection about objects created at the interface between art and science, and generation of out-of-the-box solutions to intricate problems of global relevance.

# Temporal control of cell division: switches, refractory periods and feedback control

Silvia SANTOS[1]

[1] MRC clinical sciences centre, Imperial College London, UK

## Abstract

During cell decision-making signal transduction networks dynamically change in time and space in response to cues, and thereby trigger different cellular states. The decision to divide is one of the most fundamental cellular responses and the evolutionarily conserved networks that control cell division adapt and remodel in a variety of biological contexts - during development and homeostasis, infections and malignancy, in response to drugs and stresses. A striking example of this versatility occurs during development where the same core regulators drive structurally different divisions. Divisions in the embryo are clock-like, fast, short and synchronous with no checkpoints or gap phases. With time, these divisions become longer and asynchronous. The resulting somatic like cycles have checkpoint control and gap phases, and the initiation of events is dependent on completion of early events, just like a falling domino. The question, thus, arises on how do the same cell cycle regulators self- organize and remodel in time and space to generate structurally different cell division cycles? In our lab we use human embryonic stem cells as a model for the embryonic cell cycle and monitor the activities, concentrations and spatial distribution of key cell cycle regulators in single cells, during ES cell differentiation.

In my talk I will be discussing how combining single cell imaging and omics approaches with mathematical modelling is allowing us to shed light into how cell cycle networks remodel in time and space during cell cycle transitions and during ES cell differentiation.

# Immune recognition, antagonism and phenotypic spandrel

Paul FRANÇOIS[1]

[1] Department of Physics, McGill University, Canada

## *Abstract*

Recent works in quantitative evolution have shown that biological structures are constrained by selected phenotypes in unexpected ways . This is also observed in simulations of gene network evolution, where complex realistic traits naturally appear even if they have not been explicitly selected . An important biological example is the absolute discrimination between different ligand *qualities*, such as immune decisions based on binding times to T cell receptors (TCRs) or Fc$\epsilon$RIs. I will present our evolutionary simulations in which the phenomenon of absolute discrimination is mediated by an adaptive sorting mechanim. Adaptive sorting is not achieved without detrimental ligand antagonism: a *dog in the manger* effect in which ligands unable to trigger response prevent agonists to do so. I will show indeed how absolute discrimination and ligand antagonism are interconnected. Inspired by the famous discussion by Gould and Lewontin, we thus qualify antagonism as a *phenotypic spandrel*: a phenotype existing as a necessary by-product of another phenotype. Phenotypic spandrels reveal the internal feedbacks and constraints structuring response in signalling pathways, in very similar way to symmetries structuring physical laws.

# DNA-encoded programs: In vitro models of biological networks

Yannick RONDELEZ[1]

[1] LIMMS/CNRS-IIS, Institute of Industrial Science,
The University of Tokyo

## *Abstract*

Living cells use webs of chemical reactions organized in precise networks to perform informational tasks. This concept is at the very core of systems biology, but also has link with electronic or computer science. The dynamical and information-processing properties of living cells, i.e. their ability to make decisions, sense their environment, maintain their integrity, memorize bits of information, interact, coordinate, etc. is indeed encapsulated in the topology and dynamics of their molecular networks.

I will present a method, based on standard DNA biochemistry, to build artificial analogues of cellular circuits *in test tubes*, using a simple enzymatic decoding machinery. It uses modular elementary interactions (activation, inhibition, degradation), which can be connected in tightly regulated networks of desired topology. The system is kept out of equilibrium using catalytic resolution of kinetic bottlenecks, but could also in principle be embedded in an open system. This approach was initially demonstrated by building *de novo* and *in vitro* a robust chemical oscillator: we implemented a positive and a delayed negative feedback loops, encoded in the sequence of small DNA templates, and obtained the predicted oscillatory dynamics [1]. More recently, we have extended the approach to encode other types of biological networks, such as those involved in complex ecosystems. Also, because of the simple and well-controlled environment, the chemical network is easily amenable to quantitative mathematical analysis so that the approach can be partially automatized. Other features of living systems are being integrated [2]. For example many networks, most notably morphogenic subsystems, use diffusion as a key functional tool, and I will present initial results in this direction. Linearity/non-linearity of the individual interactions is also an essential component of the function, and we have worked to make it programmable as well. Our results show that the rational cascading of standard elements opens the possibility to implement complex behaviours *in vitro*. These synthetic systems may thus accelerate our understanding of the underlying principle of biological dynamic networks and provide building blocks for the construction of more integrated emergent behaviours.

### *References*

[1] Montagne, Fujii, Rondelez, Molecular System Biology, 7:466, **2011**.

[2] Padirac *et al.* Bottom-up construction of in vitro switchable memories, PNAS **2012**.

# RNA-based circuits in vitro and in vivo

Friedrich SIMMEL[1]

[1] Systems Biophysics and Bionanotechnology, Technische Universität München, DE

## Abstract

Gene regulatory mechanisms such as anti-sense RNA, riboregulators and riboswitches, or CRISPR interference utilize base-pairing interactions and RNA conformational changes, which – compared to other biomolecular interactions – are relatively well understood and predictable. It is thus quite straightforward to artificially "cprogram" such gene regulatory processes by simply designing the proper RNA sequences. This makes RNA regulatory motifs particularly promising for applications in synthetic gene circuitry.

In the talk, we will discuss a variety of examples of synthetic RNA-based circuits, which have been recently implemented in our lab. Specifically, we will deal with a synthetic biochemical oscillator system based on simple transcriptional switches termed "genelets" (which can only be implemented *in vitro*). We will then also present a variety of rationally designed in vivo gene circuits involving so-called "toehold riboregulators" and also CRISPR-based mechanisms.

## References

[1] E. Franco, E. Friedrichs, J. Kim, R. Murray, E. Winfree and F. C. Simmel, Timing molecular motion and production with a synthetic transcriptional clock, Proc Natl Acad Sci USA 108, E784-E793 (2011)

[2] M. Weitz, J. Kim, K. Kapsner, E. Winfree, E. Franco and F. C. Simmel, Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator, Nat Chem 6, 295-302 (2014)

[3] K. Kapsner and F. C. Simmel, Partitioning Variability of a Compartmentalized In Vitro Transcriptional Thresholding Circuit, ACS Synth Biol 4, 1136-43 (2015)

# Hybrid and logical approaches for dynamical analysis of gene regulatory networks

Madalena CHAVES[1]

[1] Biocore group, INRIA Sophia Antipolis Meditérranée, FR

## *Abstract*

Mathematical modeling and control theoretic ideas play an essential role in the design and implementation of new circuits in Systems and Synthetic Biology, tasks which require a previous careful study of the dynamical behavior of the network of interactions among the different components. Since these individual biological components generally have a nonlinear behavior and present an inherent variability, mathematical models provide a first representation of the response of the synthetic cell, and this makes them obligatory for a successful implementation of the circuit.

Depending on the type of measurements available (eg., the frequency and accuracy of data points), different mathematical formalisms can be chosen to more efficiently allow a comparison between the biological and the mathematical models. Thus, in the case of less accurate or infrequent measurements, or for systems with many variables, logical models are more appropriate and can be used to great advantage. This is because logical models involve a low number of parameters and can often be analyzed with *exact* tools based on graph theory. In contrast, continuous models of differential equations generate more detailed representations of a system but also require much more data and information for parameter estimation and model validation. We will also introduce a third *intermediate* class of models which provides a continuous representation of trajectories while using discrete *activation* functions: these are known as piecewise affine (PWA) models, since they can be described as a concatenation of linear systems of differential equations.

Several examples will be used throughout the presentation to illustrate applications of the different mathematical formalisms, such as some simple regulatory motifs, a model for the circadian rhythm in cyanobacteria, or a model for growth rate in *E. coli*.

## *References*

[1] G. Bernot, J.-P. Comet, A. Richard, M. Chaves, J.-L. Gouzé, and F. Dayan. *Modeling and analysis of gene regulatory networks*. In "Modeling in Computational Biology and Biomedicine", F. Cazals and P. Kornprobst

Eds, Springer-Verlag Heidelberg (2013), pp. 47-80 (available for download from Springer's website, as Sample pages).

[2] M. Chaves and M. Preto. *Hierarchy of models: from qualitative to quantitative analysis of circadian rhythms in cyanobacteria*, Chaos (Focus issue), 23(2), pp. 025113, 2013.

[3] M. Chaves, E. Farcot, and J.-L. Gouzé. *Probabilistic approach for predicting periodic orbits in piecewise affine differential models*, Bull. Math. Biology, 75(6), pp. 967-987, 2013.

# G-Networks for Gene Regulatory Networks

Erol GELENBE[1]

[1] ISN, Imperial College London, UK

## Abstract

G-Networks are an infinite state-space representation of Gene Regulatory Networks where a variable concentration of different genetic agents interact to enhance or inhibit each other's action and affect the concentration of the different genetic agents. We will describe the model and derive its dynamics and equilibrium behaviour. G-Networks also have efficient parameter identification algorithms which are of $O(n^3)$ complexity for n agents. Based on these tools we will show how they can be used to exploit data so as to detect anomalies that can lead to disease.

## Brief biography

Erol Gelenbe is the Gabor Professor at Imperial College and the Head of Intelligent Systems and Networks. For his interdisciplinary research covering computer science and other areas, he was elected a Fellow of the Academies of Science of Belgium, Hungary, Poland and Turkey, and of the French National Academies of Technology. He was awarded Chevalier de l'Ordre national de la Legion d'honneur, and several other prizes in France, the USA, Italy, Hungary and Turkey.

# Design and synthesis of a minimal bacterial genome

John GLASS[1]

[1] Synthetic Biology and Bioenergy Group, JCVI, Rockville, US

## *Abstract*

Whole-genome design and complete chemical synthesis were used to minimize the 1079-kb synthetic genome of Mycoplasma mycoides JCVI-syn1.0. An initial design, based on collective knowledge of molecular biology plus limited transposon mutagenesis data, failed to produce a viable cell. Improved transposon mutagenesis methods revealed a class of quasi-essential genes needed for robust growth, explaining the failure of our initial design. Three cycles of design, synthesis, and testing, with retention of quasi-essential genes, produced JCVI-syn3.0 (531kb, 473 genes), with a genome smaller than that of any autonomously replicating cell found in nature. JCVI-syn3.0 retains almost all genes involved in synthesis and processing of macromolecules. Surprisingly, it also contains 149 genes with unknown biological functions. JCVI-syn3.0 is a versatile platform for investigating the core functions of life, and for exploring whole-genome design.

# The modular cell in synthetic biology: its structural and functional expansion with connection to the system level

Stefan SCHILLER[1]

[1] Zentrum für Biosystemanalyse (ZBSA) and Freiburg Institute for Advanced Studies (FRIAS), 79104 Freiburg, Germany

## *Abstract*

Life is embedded into a fascinating molecular complexity of dynamic structure & function forming processes we aim to understand and to control. The synthetic biology approach we use interprets synthetic biology as transformation and organization of molecules and molecular systems within complex systems creating new system properties structuring and functionalizing space. Thus, the synthetic biology approach focuses on embedding new cross-compatible molecules & molecular networks as structural and functional modules for the functional expansion of the cell.

Molecular operations take place at a dimension where supramolecular assemblies create properties not present at the level of individual molecules nor at the meso- or macroscale. Thus, this field – termed nanotechnology – gained some interest especially due to new (interfacial) material properties based on structural and functional features at the nanoscale, the scale of large molecular assemblies. Biology, especially cell and molecular biology is inherently dealing with this dimension and its effects. In contrast to the aforementioned area biological systems have, beside others, two important features: dynamic behavior and the creation of structures and their reorganization upon the dissipation of energy. Dynamic compatibility is one important requirement for our system-module properties we only recently began to realize as important factor. In this context we will consider time scales for dynamic processes in the formation and change of space-structuring and, thus function generating molecular architectures e.g. assembling and tuning enzyme cascades.

In order to realize such features these systems require a certain level of complexity regarding the number and identity/(multi)functionality of their components. Thus, theories and approaches which represent biological parts in analogy to classical engineering with static (molecular) structures and building blocks already describe the modules constituting such systems in a very limited and oversimplified way. Especially our insufficient estimates (intuitive errors) for dynamic effects in complex systems can be a fundamental limitation on our way to efficiently expand, change or rebuild cellular systems. Approaches to reflect on these pitfalls in order to suppress some of these limitations in our

"mental" approach to think and control dynamics and complexity can be found in a set of interesting experiments demonstrating these mistakes & failures [1]. The awareness of our mind-models, while we are thinking about and dealing with complex and non-linear systems, is of great importance to efficiently deal and to improve biological systems.

The correlation of analyzing complex systems as stepwise thinking & deduction as scientific tool, with the physical world, mirror images the discrete nature in our current quantum oriented thinking. Discreteness as such, is not only reflected in the molecular description of matter and its energy levels, but recently also reflected in mathematical descriptions such as noncommutative geometry. Why noncommutative geometry [2] might be important in synthetic & systems biology? Well, the non-identical non-interchangeable space and the spatiotemporal effects in structure formation, signal/information processing... are currently neither handled experimentally nor theoretically – but we know they are essential in cellular processes. With a current emphasis on quantum dynamics, molecular interactions, fractal dimensions & chaotic systems we face some interesting theoretical elements in our thinking. Coherence in quantum phenomena/interaction and selection rules is for instance recently discussed to be important in microtubule assembly [3, 4]. Beside the aforementioned areas especially fractal dimension, known from the fractal geometry of B. Mandelbrot allow for important cellular features such as a high level of organization, shape irregularity, fractal dimension, as well as functional and morphological self-similarity... many of them important e.g. to understand cellular organelles.

These are the fundamental questions which are essential to be reflected on and to looked after how they shape our ideas and approaches and how we can think new correlations, how to identify new ways to assemble systems, to exchange energy and information under dynamic, complex, crowded and energy dissipative nonlinear and fractal (broken dimensions) regimes.

After these theoretical reflections on important elements in synthetic biology, and science in general, molecular ways towards new cellular modules allowing to expand cellular functions, will be highlighted.

Taking a functional engineering approach we design a toolbox of biogenic tectons (tecton = architectural building block) on the DNA and protein level delivering these tectons as cellular building blocks and networks with regulatory and structural functions, which did not previously exist in nature [5, 6]. The novel format we developed was termed one-vector toolbox platform (OVTP) and allows for the assembly of highly repetitive and asymmetric genes (Figure 1). Realizing such protein-based modular biological devices resembling a close systematic link to the so-called BioBricks, we de-

sign biogenic protein tectons within cells showing complex dynamic structure forming behavior, such as adjustable structured self-assembly of organelle-like NanoBioSystems [7, 8].
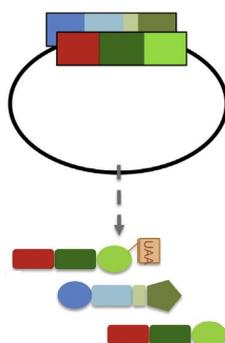


**Figure 1**:   The one-vector toolbox platform (OVPT) allows to assemble complex genes and to code for new chemical functionalities via an expanded genetic code.

The building blocks used are defined assemblies of pentapeptide sequences (VPGXG) where "X" can be any amino acid beside proline, derived from tropoelastin partial sequences, thus called elastine-like proteins (ELPs). The adjustment of homo or hetero-blocks with e.g. hydrophilic or hydrophobic amino acids at position "X" allows for combinatorial complexity creating various compartment-like structures inside the cell (Figure 2).
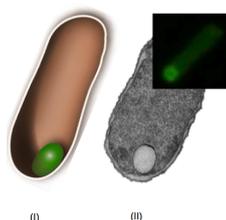


**Figure 2**:   Cellular compartments shown as model, electron microscopy image and fluorescent image.

Current applications include the functionalization of these compartments with unnatural amino acids further allowing to implement new chemical functions, the coexpression of proteins with membrane anchors and the covalent protein-protein coupling onto these compartments allowing first steps towards the assembly of enzyme cascades.

These efforts are complemented by redesigning the translational networks of the cell expanding the functionality of the genetic code with xenobiotic

elements such as unnatural amino acids further increasing the functional space of defined molecular protein modifications combining chemical and biosynthetic methods. An interesting approach in this context is the transformation of cofactors to prosthetic groups within enzymes and approaches to control the assembly of proteins towards enzyme cascades via protein switches and optogentic tools. Here we envision to functionally expand the chemical space of enzymatic transformations to be combined with intracellular *de novo* organelles.

By designing interacting protein architectures and implementing novel interaction systems on the nanoscale via bioorthogonally encoded chemical entities/signals we emulate the coexistence of several elements allowing for a functional expansion of the cell by molecular system design of our functional modules. Implementing the controlled assembly of organic-inorganic nano-architectures we termed "protein adaptor-based nanoobject assembly" (PAB-NOA) [9] opens up new roads towards functional materials with new properties with molecular precision for possible applications in modulating molecular quantum phenomena e.g. in energy converting or harvesting systems *in vitro* and *in vivo*. This approach is an interesting modular feature which may allow to expand photosynthetic reactions far away from the classical photosynthesis. Such approaches do not only yield new materials and small molecules, but may allow different routes to implement the biological energy and mass flow under sustainable aspects. The scalability of the molecular role for the transformation of matter and the conversion of energy in the context of *in vivo* and *in vitro* applications and their embedding into the natural material and energy cycle allows for a vast range of applications especially in the context of a sustainable bioeconomy, exemplified in the picture below.
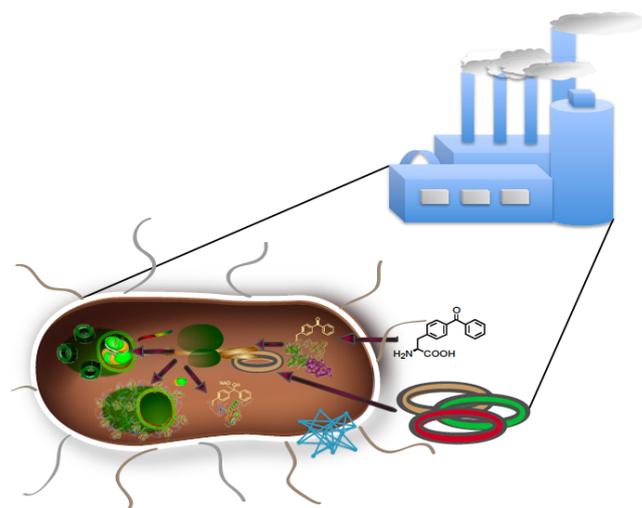
## References

[1] Dietrich Dörner, *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations* (1996) Published by Basic Books; New York. ISBN 9780201479485

[2] Connes, Alain; Marcolli, Matilde (2008), "A walk in the noncommutative garden", An invitation to noncommutative geometry, World Sci. Publ., Hackensack, NJ, pp. 1–128, arXiv:math/0601054

[3] S. Hagan, S. R. Hameroff, and J. A. Tuszyński "Quantum computation in brain microtubules: Decoherence and biological feasibility" *Phys. Rev.* E 65, 061901 - Published 10 June 2002

[4] Stuart Hameroff & Roger Penrose, In: *Toward a Science of Consciousness - The First Tucson Discussions and Debates*, eds. Hameroff, S.R., Kaszniak, A.W. and Scott, A.C., Cambridge, MA: MIT Press, pp. 507–540 (1996)

[5] Huber, M., Schreiber, A., Benz, K., Schiller, S. M. "Introducing a combinatorial DNA-toolbox platform constituting defined protein-based biohybrid materials" *Biomaterials* 2014, 35, 31, 8767-79

[6] Schiller, Stefan M., "Protein Tectons in Synthetic Biology: The expansion of cellular functionality combining chemical biology of small organic molecules with protein tectons – unnatural amino acids, protein based biohybrid materials & de novo organelles", in *Synthetic Biology*, ed. Giese, B., von Gleich, A., Pade, C., Wigger, H. (Springer, 2014, ISBN 978-3-319-02782-1)

[7] Huber, M. C., Schreiber, A., von Olshausen, P., Varga, B. R., Kretz, O., Joch, B., Barnert, S., Schubert, R., Eimer, S., Kele, P., Schiller, S. M. "Designer amphiphilic proteins as building blocks for the intracellular formation of organelle-like compartments" *Nature Materials* 2015, 14(1), 125–132

[8] Schreiber, A., Schiller, S. M "Nanobiotechnology of Protein Compartments: Steps towards Nanofactories" *Bioinspired, Biomimetic and Nanobiomaterials* (ICE Virtual Library) 2013, 4, 2, 157–164

[9] Schreiber, A., Huber, M. C., Cölfen, H., Schiller, S. M. "Protein Adaptor with Genetically Encoded Interaction-Sites Guiding the Hierarchical Assembly of Plasmonically Active Nanoarchitectures" , DOI: 10.1038/ncomms7705 *Nature Communications*, 2015, 6, article number 6705

# Human-Scale Metabolic Network of Central Carbon Metabolism. Application to serine metabolism from glutamine in Cancer Cells.

Bertrand Beauvoit[1], Sophie Colombié[1], Razanne Issa[2], Jean-Pierre Mazat[2], Christine Nazaret[3], Sabine Pérès[4] (alphabetical order)

[1] INRA, UMR 1332 Biologie du Fruit et Pathologie, F33883 Villenave d'Ornon, France and Univ. Bordeaux 146 rue Léo-Saignat, F 33076 Bordeaux Cedex France

[2] IBGC CNRS UMR 5095 & Universite de Bordeaux, 1, rue Camille Saint-Saëns 33077 Bordeaux cedex, France

[3] Institut de Mathématiques de Bordeaux, ENSTBB-Institut Polytechnique de Bordeaux

[4] Laboratoire de Recherche en Informatique (LRI) Bât 650 Université Paris-Sud 91405 Orsay Cedex France and MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France

## *Abstract - Introduction*

Central carbon metabolism (glycolysis, gluconeogenesis, pentose phosphate pathway, fatty acid synthesis, TCA cycle) is an important part of metabolism both for synthesis of new molecules (amino acids, nucleotides, fatty acids etc.) but also from the energy point of view (ATP synthesis) and the redox balance of the cell (NAD/NADH and NADP/NADPH). The knowledge of genomes leads to the construction of genome-scale models (GSM) involving all the enzymes possibly encoded in the genome. The best example of such a model is Recon2 involving 7,440 reactions and 5,063 metabolites [1]. Due to their big size, it is difficult to study these models and the only possible approach is Flux Balance Analysis (FBA) looking for flux values able, at steady-state, to maximize or minimize some objective function. The calculation of all the Elementary Flux Modes (EFMs), i.e. the minimal pathways inside the metabolic network at steady-state are out of the possibilities of our computers and the lack of knowledge of the amount of most of the enzymes and their kinetic properties prevent to develop large representative dynamical systems. Furthermore it is difficult on such big systems to understand their functioning in special situations (normal or pathological) and even to be sure that the results obtained are not artefactual or biased.

For all these reasons we decided to develop simpler models still representing the main architecture of the whole metabolism but with fewer simulated reactions which are aggregations of the actual reactions. The advantage of such models is to be more easily tractable and more understandable. Furthermore they can be approached with a greater panel of methods such as analysis of

EFMs, FBA and FVA. In addition, their dynamical behavior can be studied with some reasonable hypotheses on their kinetic laws.

We will systematically apply, in a logical order, several theoretical approaches for metabolic networks as we already did in [2]. We will emphasize the advantages and the drawbacks of each approach and show how we can use all of them to gain a better understanding of the behavior of a metabolic network.

## 1    Description of a Human-Scale metabolic (HSM) model of central carbon metabolism

The reactions involved in HSM13 are listed in Appendix 2. They consist in a simple version of the Krebs cycle already described in [2] (reactions PDH, CS, K4, K567, MDH2 and PYC with the addition of glutamate dehydrogenase (GLUD1), in the glycolysis summarized in 5 steps, G1 (hexokinase + phosphoglucose isomerase), G2 (phosphofructokinase + aldolase + triose-phosphate isomerase), G3 (glyceraldehyde-3P dehydrogenase + phosphoglycerate kinase), ENOMUT (enolase + phosphoglycerate mutase) and PK (pyruvate kinase) extended by the reversible LDH (lactate dehydrogenase). The gluconeogenesis consists in the reversible reactions of glycolysis with PEPCK1, GG3 (triose phosphate isomerase + aldolase + fructose-1,6-biphosphatase) and GG4 (phosphogluco isomerase + glucose-6-phosphatase). There is also a mitochondrial PEPCK named PEPCK2.

The reactions of pentose phosphate pathway (PPP) are summarized in PP1 (oxidative part of PPP) and PP2 (non-oxidative part of PPP).
The synthesis of nucleotide bases is represented by a simplification of purine biosynthesis PUR, which is important here because it is an entry of glutamine in metabolism and its conversion to glutamate. The other entry of glutamine is in the mitochondria through transporter T8. It will give glutamate in mitochondria through the operation of glutaminase GLS1. The synthesis of serine from 3-phosphoglycerate involves 3 steps: a dehydrogenase, a transaminase involving the glutamate/2-oxoglutarate couple and a phosphatase. The three steps are assembled in one reaction SERSYNT.

The malate/aspartate shuttle (MAS) is fully represented with its two exchangers, the malate/2-oxoglutarate exchanger T2 (OGC) and the glutamate / aspartate exchanger T4 (GLAST), the malate dehydrogenases (cytosolic MDH1 and mitochondrial MDH2) and the glutamate- oxaloacetate transaminases (cytosolic GOT1 and mitochondrial GOT2). A detailed representation of MAS was necessary because the MAS enzymes are not always used with the stoichiometry of MAS for exchange of NADHc for NADHm i.e. MAS components are not always used to run the MAS as such.

**Figure 1**: HSM13-3 metabolic network. The abbreviations are given in appendix A.

The synthesis of fatty acids starts with citrate lyase (CL) and is represented in the case of palmitate by the reaction PL1. Finally, the oxidative phosphorylations are represented by RC (respiratory chain) AS (ATP synthase), ANT, the ADP/ATP exchanger, T5 the Pi carrier and L the membrane leak of protons. ATP usage is symbolized by ATPASE. Because all the metabolites involved in these last reactions are considered as external, these reactions are not actually involved in the model.

Several entries are considered: entry of glucose (GLUCUP), of glutamine (GLNUP) and of aspartate (ASPUP). The outputs are serine (SEROUT), bases (BASES) and palmitate (Palmitate). LACIO symbolizes the possible input / output of lactate in cytosol.

HSM13 Model contains 27 Internal metabolites (constrained to steady-state), 23 external metabolites that do not participate to steady-state and 50 reactions (24 reversible and 26 irreversible).

## 2 The Elementary Flux Modes (EFM) of the model leading to Serine biosynthesis as a function of Carbon entry

### Introduction

An **elementary flux mode (EFM)** [3, 4] is a minimal set of enzymes that can operate at steady state with all irreversible reactions used in the appropriate direction. All flux distributions in the living cell are non- negative linear combinations of elementary modes. The decomposition is not necessarily unique.

A related concept was defined by the group of Palsson: **Extreme pathway** [5] in which every reversible internal reaction is split in two irreversible reactions (one is the forward reaction, the other is the reverse reaction). The number of EFMs is finite but can be great. Their comprehensive description gives all possibilities to browse the metabolic network. As we will see below, some of them are not trivial. Considering EFM is useful because they represent extreme simplest situations of metabolic pathways.

The analysis of HSM13 with metatool [6] gives 7018 EFMs among which 3978 lead to serine biosynthesis. We will focus our analysis on those EFMs leading to serine biosynthesis. They will be classified according to the carbon entry in the metabolic network: aspartate alone, glucose alone, glutamine alone (via purine pathway or via mitochondria entry), glutamine + glucose. From a general point of view there are several possible pathways for serine biosynthesis. In all cases, serine synthesis requires 3PG (3 phosphoglycerate) as a precursor and a transamination reaction involving the conversion of glutamate to 2-oxoglutarate (AKG). It is thus necessary either to have a recycling of AKG to glutamate or to have a continuous glutamate synthesis (from glutamine entry for instance with an output of AKG or of one of its derived products).

**Serine biosynthesis from Aspartate (Fig. 2)** Aspartate is a good precursor of glutamate through the activity of aspartate aminotransferase (GOT1 gene in cytosol or GOT2 gene in mitochondria). 1011 EFMs with aspartate entry (ASPUP) are counted in the 3978 EFM involving a serine output (SEROUT). Among these 1011 SEROUT-ASPUP 640 also include an input of glucose
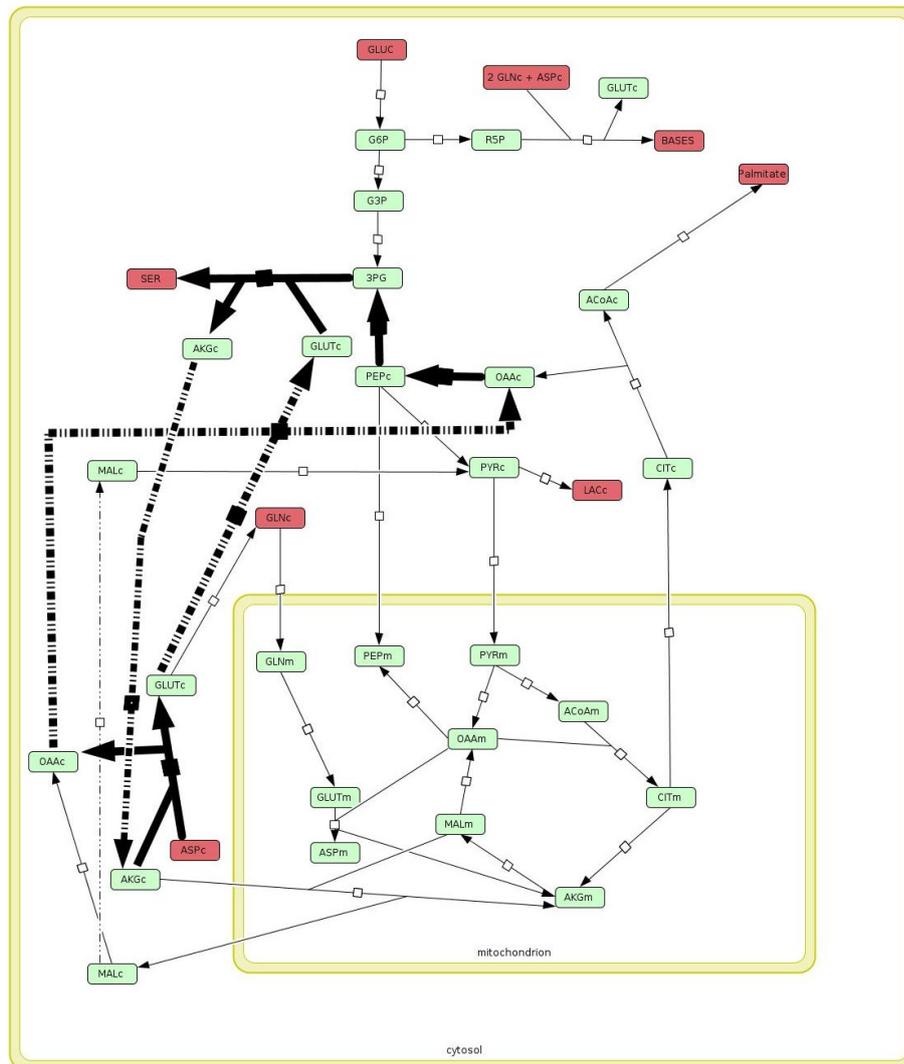
**Figure 2**:  Simplified representation of serine biosynthesis from aspartate in cytosol (EFM 5141)

GLUCUP. The shortest pathway for serine biosynthesis is observed with asparte input (EFM 5141 involving 6 reactions). Aspartate gives glutamate and the AKG produced in serine biosynthesis is recycled in OAA by GOT1 which gives PEP (PEPCK1) and then 3PG. The 4 carbon atoms of aspartate are transformed in the 3 carbon atoms of serine and in $CO_2$ (PEPCK1).  The maximal yield is **1 Asp => 1 Ser**.  The common motif ASPUP PEPCK1 SEROUT SERSYNT ENOMUT emerges from all these EFM.

The GOT reaction followed by PEPCK may occur either in cytosol (EFM 5141) or in mitochondria (EFM 5126) or both (EFM 5127 and 5140). In the case of a mitochondrial GOT-PEPCK (EFM 5126), some of the transporters of the Mal/Asp shuttle are used in the reverse direction (it does not mean that the Mal/Asp shuttle is reversed, only some transporters are used in the reverse direction). In addition, the exchanger Pi/Mal is necessary to recycle the Mal involved in the Mal/AKG exchanger.



**Figure 3**: Simplified representation of serine biosynthesis from glucose alone (EFM 1047)

***Serine biosynthesis from glucose alone (Fig. 3)***   We get 2669 EFMs
SEROUT-GLUCUP and among these, 2500 SEROUT-GLUCUP- GLNUP. The
most obvious pathway to serine synthesis from glucose alone is with glycolysis
synthesizing 3PG. The AKG produced in serine synthesis has to be recycled by
glutamate dehydrogenase activity (in the reverse reaction) in the mitochondria.
In this process the 6 carbon atoms of glucose gives 2 molecules of serine
(3C) and one NH3 is consumed. The maximal yield is: **1 Gluc + NH3 => 2
Ser**. The shortest EFMs is EFM 1047 (10 reactions). Many variations around
the common motif G1 G2 GLUCUP (2 SEROUT) (2 SERSYNT) (2 G3) (-2
GLUD1) exist (see below the ACoM analysis)

***Serine biosynthesis from glutamine (Fig. 4 and 5)***
3726 SEROUT-GLNUP are sorted from the SEROUT EFMs and among these,
2500 also involve an input of glucose (SEROUT-GLNUP-GLUCUP). The len-
gth of these EFMs is greater than the one of the above EFM indicating a more
tortuous pathway for serine synthesis from glutamine.

Glutamine can give glutamate in two ways: either by glutaminase (GLS1)
in mitochondria or in cytosol via purine biosynthesis. In all cases the AKG
formed in serine synthesis is converted to OAA (-Got1 for instance but also by
many other pathways) and then to PEP (PEPCK) and 3PG

***-Serine biosynthesis with glutaminase in mitochondria alone***   EFM 72,
is the shortest such EFM with 11 reactions (see Fig. 4) but also EFMs 52,
73 and 1615 with 13 reactions) evidencing the common motif GLNUP GLS1
K567 PEPCK2 SEROUT SERSYNT ENOMUT MDH2 -T9. 3PG for serine
synthesis comes from AKG (=> OAA => PEPc). The maximal yield is **1 Gln
=> 1 Ser**.

***-Serine biosynthesis via purine biosynthesis in cytosol alone***   Among
the 3726 EFMs leading to serine from glutamine, 3273 involve the PUR reac-
tion. Among them, 876 EFMs use T8 indicating an additional glutamine entry
in mitochondria and, correspondingly, the remaining 2397 EFMs incorporate
glutamine through the PUR reaction only. Among these, EFM 816 and EFM
5463 are the shortest with 13 reactions (Fig. 5). The PUR aspartate is either
furnished by a direct entry (ASPUP in EFM 5463) or synthesized by GOT 1
from one of the two glutamate produced in PUR (EFM 816). In this EFM, the
oxidative pentose phosphate pathway is used to make R5P from glucose.

Most of the EFMs involving PUR reaction involve also a direct input of
glutamine through GLNUP. Only three do not, EFMs 5214, 5217 and 5402. In
these cases, their necessary glutamine is synthesized from glutamate through
glutamine synthase (GLS1) with aspartate and lactate uptake.
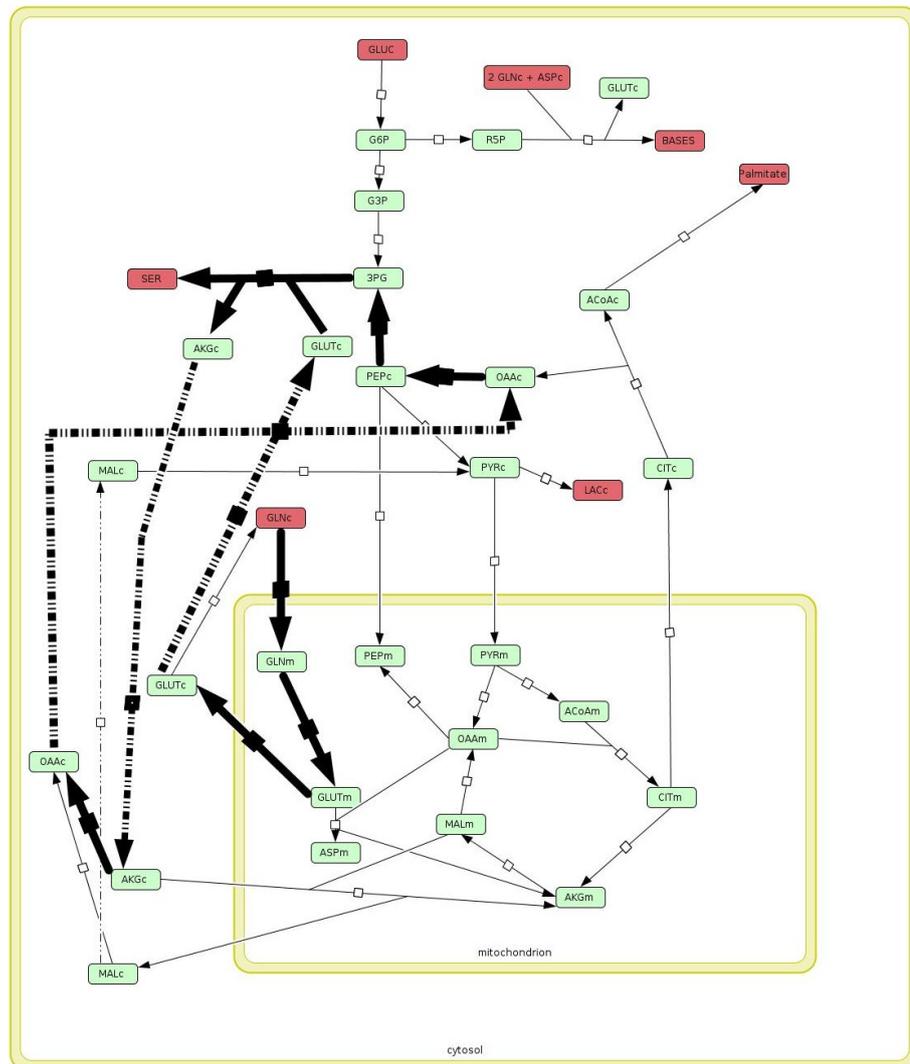
**Figure 4**:   Simplified representation of serine biosynthesis via glutamine entry in mitochondria.

**Figure 5**:   Simplified representation of serine synthesis from glutamine uptake in purine biosynthesis.

Note that in all these cases with purine synthesis, it is necessary to form 3PG for serine biosynthesis but also G3P and G6P to R5P synthesis.

Several other EFMs could be described. They are generally more complex with several entries and outputs. Their classification has be done using ACoM.

### 3 ACoM classification of the Elementary Flux Modes (EFM) of HSM13 leading to Serine biosynthesis

ACoM, is a bi-clustering method especially developed for the classification of the huge number of EFMs [7]. The application of ACoM to the set of the 3978 EFMs leading to serine biosynthesis leads to 13 common motifs of length greater than 8 with a similarity $\geq 12$ (Table 1). We find the common motifs already mentioned but also some others. We note the absence of motif involving ASPUP, probably because these motifs and the EFMs involving this reaction are too short. We get several motifs with glucose entry, glutamine entry in mitochondria or in PUR reaction, but also several motifs with entry of lactate.

The importance of PEPCK1 or PEPCK2 must be stressed. These reactions appear in 5 common motifs on 13.

| COMMON MOTIF | SIZE | # EFMs | Phenotype – Biological Meaning |
|---|---|---|---|
| PEPCK2 SEROUT SERSYNT T6 ENOMUT - LACIO - LDH - T7 | 8 | 412 | Serine synthesis from lactate via PEPCK2 |
| PEPCK2 SEROUT SERSYNT ENOMUT - GLUD1 - LACIO - LDH - T7 - T9 | 9 | 43 | Serine synthesis from lactate via PEPCK2 |
| PEPCK1 PYC SEROUT SERSYNT T6 ENOMUT - GLUD1 - LACIO - LDH MDH1 T2 - T9 | 12 | 14 | Serine synthesis from lactate via PEPCK1 |
| SEROUT SERSYNT ENOMUT GOT2 - LACIO - LDH T4 - T9 | 8 | 59 | Serine synthesis from lactate |
| GLNUP GLS1 SEROUT SERSYNT - GOT1 GOT2 T4 T8 | 8 | 327 | Serine synthesis from glutamine in mitochondria |
| G1 GLUCUP PP1 SEROUT SERSYNT T6 G3 PP2 - T9 | 9 | 157 | Serine synthesis from glucose, part of glycolysis and PPP |
| G1 GLNUP GLUCUP K567 PUR SEROUT SERSYNT T2 | 8 | 1094 | Serine synthesis fromglucose and glutamine entry in PUR |
| G1 GLUCUP SEROUT SERSYNT G3 - GOT1 GOT2 T4 | 8 | 336 | Serine synthesis from glucose |
| G1 GLUCUP PYC SEROUT SERSYNT T6 G3 - GLUD1 T2 - T9 | 10 | 162 | Serine synthesis from glucose |
| CL GLNUP GLS1 PL1 SEROUT SERSYNT T1 T8 | 8 | 968 | Serine and FA synthesis from glutamine in mitochondria |
| PEPCK2 SEROUT SERSYNT T6 ENOMUT - GOT2 ME1 T2 - T4 - T7 | 10 | 145 | Serine synthesis via PEPCK2 |
| PEPCK1 PEPCK2 SEROUT SERSYNT - GOT2 T2 - T4 - T7 | 8 | 70 | Serine synthesis via PEPCK1 and PEPCK2 |
| G1 GLUCUP PP1 SEROUT SERSYNT G3 LACIO LDH PP2 | 9 | 148 | Serine synthesis from glucose, part of glycolysis, PPP and lactate output |

**Table 1**: Common motifs of length $\geq 8$ (and similarity $\geq 12$) calculated with ACoM. The table describes successively the common motif, its size, the number of EFMs including this common motif and their biological meaning.

## 4 Flux-Balances-Analysis (FBA) and Flux-Variability-Analysis (FVA) for serine biosynthesis from glutamine

Flux Balance Analysis (FBA) is a method developed by the group of Palsson [8] aiming at optimizing the flux values in a metabolic network to fulfil a peculiar objective such as cell growth or ATP production for instance. The objective is formalized under the form of an "objective function" (a rate equation of ATP consumption in the case of optimizing ATP production). Known constraints on the fluxes can be added such as minimal and maximal values (otherwise maximizing a flux will lead to infinity). FBA can be applied in many other contexts to analyze the phenotypes and capabilities of organisms upon different environmental and genetic perturbations (KO genes for instance). The optimization is directed at the metabolic fluxes values, without any knowledge of the underlying rate functions so that it can be applied to big genome scale networks for which not all steps are known in detail. We will apply this approach to our simple model with maximization of serine output as objective function. For that we use FAME software [9]. Note that in general the solution of the optimization is not unique. Several values of different flux can lead to the same optimized value. Flux-Variability- Analysis (FVA) gives the variation of each variable flux giving the same optimized value. Adding the constraint of minimizing all the fluxes, gives a unique solution which is in general a simple one. In all cases we will impose this last constraint of minimal fluxes.

### 4.1 Maximal flux of serine synthesis in the case of aspartate supply

Applying FAME with the constraint $0 < ASPUP < 1$ and the other entry fluxes (GLUCUP, GLNUP, LACIO) equal to zero, we obtain the following result in maximizing the reaction SEROUT:

| Reaction ID | Flux value | Lower bound | Upper bound | Reduced costs (scaled RC) | Reaction info (corrected for flux direction) |
|---|---|---|---|---|---|
| ASPUP | 1 (1) | 0 | 1 | 0.0 (0) | ASPUP - 1 ASP ==> 1 ASPc |
| ENOMUT | 1 | -50 | 50 | 0.0 (0) | ENOMUT - 1 PEPc ==> 1 s_3PG |
| GOT1 | -1 | -50 | 50 | 0.0 (0) | 1 ASPc 1 AKGc ==> GOT1 - 1 GLUTc 1 OAAc |
| PEPCK1 | 1 | 0 | 50 | 0.0 (0) | PEPCK1 - 1 OAAc 1 ATPc ==> 1 PEPc 1 ADPc |
| SEROUT | 1 | 0 | 50 | 0.0 (0) | SEROUT - 1 SERc ==> 1 SER |
| SERSYNT | 1 | 0 | 50 | 0.0 (0) | SERSYNT - 1 s_3PG 1 GLUTc 1 NADc ==> 1 SERc 1 AKGc 1 NADHc |

**Table 2**: Minimal fluxes maximizing serine synthesis from aspartate at steady-state

which is exactly the EFM 5141 of Fig. 2, meaning that EFM 5141 is not only the simplest EFM leading to serine, but also an EFM with the maximal yield serine/aspartate.

We will continue exploring the synthesis of serine from other supply such as glucose and glutamine looking for minimal flux still maximizing the reaction SEROUT.

### 4.2   Maximal flux of serine synthesis in the case of glucose supply

In this case we fix the entry of glucose (GLUCUP = 1) and the other entry ASPUP = GLNUP = LACIO = 0. The minimal fluxes giving the maximum of serine synthesis is represented below and corresponds to EFM 1047:

| Reaction ID | Flux value | Lower bound | Upper bound | Reduced costs (scaled RC) | Reaction info (corrected for flux direction) |
|---|---|---|---|---|---|
| G1 | 1 | 0 | 100 | 0.0 (0) | G1 - 1 GLUCc 1 ATPc ==> 1 G6P 1 ADPc |
| G2 | 1 | 0 | 100 | 0.0 (0) | G2 - 1 G6P 1 ATPc ==> 2 G3P 1 ADPc |
| G3 | 2 | -100 | 100 | 0.0 (0) | G3 - 1 G3P 1 NADc 1 ADPc 1 Pic ==> 1 s_3PG 1 NADHc 1 ATPc |
| GLUCUP | 1 (1) | 0 | 1 | 0.0 (0) | GLUCUP - 1 GLUC ==> 1 GLUCc |
| GLUD1 | -2 | -100 | 100 | 0.0 (0) | 1 AKGm 1 NADHm 1 NH3 ==> GLUD1 - 1 GLUTm 1 NADm |
| SEROUT | 2 | 0 | 100 | 0.0 (0) | SEROUT - 1 SERc ==> 1 SER |
| SERSYNT | 2 | 0 | 100 | 0.0 (0) | SERSYNT - 1 s_3PG 1 GLUTc 1 NADc ==> 1 SERc 1 AKGc 1 NADHc |
| T2 | 2 | -100 | 100 | 0.0 (0) | T2 - 1 AKGc 1 MALm ==> 1 AKGm 1 MALc |
| T3 | -2 | -100 | 100 | 0.0 (0) | 1 MALc 1 Pim ==> T3 - 1 MALm 1 Pic |
| T9 | -2 | -100 | 100 | 0.0 (0) | 1 GLUTm 1 Hm ==> T9 - 1 GLUTc 1 Hc |

**Table 3**:  Minimal fluxes maximizing serine synthesis from glucose at steady-state

### 4.3   Maximal flux of serine in the case of glutamine supply

In this case we fix the entry of glutamine (GLNUP = 1) and the other entry ASPUP = GLUCUP = LACIO = 0. The minimal fluxes giving the maximum of serine synthesis is represented below and corresponds to EFM 72 (see Table 4)

### 4.4   The effect of PEPCK on serine biosynthesis

We noticed the frequent occurrence of PEPCK 1 or 2 in the EFMs leading to serine. There are also some experimental evidence that PEPCK regulates the central carbon metabolism [10]. In order to check this point, we study the effect of decreasing the PEPCK1 activity (with PEPCK2 = 0) on the maximization of serine biosynthesis using FAME. The result is represented in Fig. 6 which shows an inversely proportional decrease in serine synthesis with the inhibition of PEPCK1. A transient accumulation of palmitate is evidenced.

| Reaction ID | Flux value | Lower bound | Upper bound | Reduced costs (scaled RC) | Reaction info (corrected for flux direction) |
|---|---|---|---|---|---|
| ENOMUT | 6 | -100 | 100 | 0.0 (0) | ENOMUT - 1 PEPc ==> 1 s_3PG |
| GLNUP | 6 (6) | 6 | 6 | 4.0 (4) | GLNUP - 1 GLN ==> 1 GLNc |
| GLS1 | 6 | 0 | 100 | 0.0 (0) | GLS1 - 1 GLNm ==> 1 GLUTm 1 NH3 |
| K567 | 6 | 0 | 100 | | K567 - 1 AKGm 1 NADm 1 Pim 1 ADPm ==> 1 MALm 1 NADHm 1 CO2 1 ATPm |
| MDH1 | 6 | -100 | 100 | 0.0 (0) | MDH1 - 1 MALc 1 NADc ==> 1 OAAc 1 NADHc |
| PEPCK1 | 6 | 0 | 100 | 0.0 (0) | PEPCK1 - 1 OAAc 1 ATPc ==> 1 PEPc 1 ADPc |
| SEROUT | 6 | 0 | 100 | 0.0 (0) | SEROUT - 1 SERc ==> 1 SER |
| SERSYNT | 6 | 0 | 100 | | SERSYNT - 1 s_3PG 1 GLUTc 1 NADc ==> 1 SERc 1 AKGc 1 NADHc |
| T2 | 6 | -100 | 100 | 0.0 (0) | T2 - 1 AKGc 1 MALm ==> 1 AKGm 1 MALc |
| T8 | 6 | -100 | 100 | 0.0 (0) | T8 - 1 GLNc ==> 1 GLNm |
| T9 | -6 | -100 | 100 | 0.0 (0) | 1 GLUTm 1 Hm ==> T9 - 1 GLUTc 1 Hc |

**Table 4**: Minimal fluxes maximizing serine synthesis from glutamine at steady-state
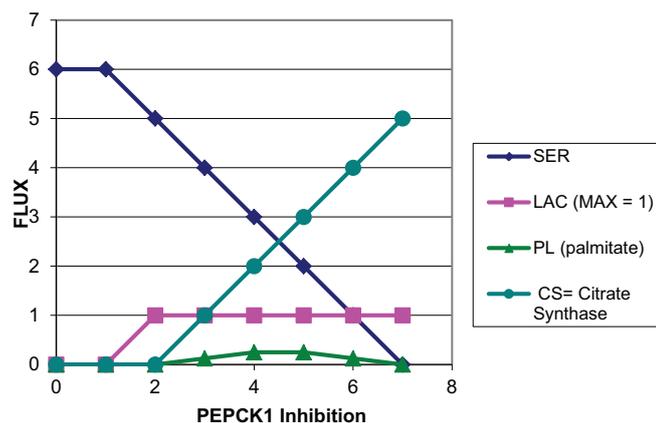


**Figure 6**: The glutamine entry is fixed at 6 (no glucose nor aspartate entry). The maximal activity of PEPCK 1 is 7 (PEPCK2 = 0). The PEPCK1 inhibition leads to a parallel inhibition of serine production (dark blue curve) since the value of 6. The carbon accumulation due to glutamine entry fixed at 6 is as a matter of priority eliminated in lactate output (pink). Because lactate output is limited to 1, the extra carbons are then dissipated in CO2 in the Krebs cycle (turquoise). A transient palmitate production (green) occurs.

## 5   A dynamical (ODE) approach of the serine biosynthesis with HSM13

(COPASI `http://www.copasi.org` ;
Berkeley Madonna `http://www.berkeleymadonna.com`).

In the case of the simplest synthesis of serine from asparte (EFM 5141), the dynamical system of our metabolic network can be written:

```
d/dt(SER) = vSEROUT
d/dt (SERc) = vSERSYNT - vSEROUT
d/dt (s3PG) = vENOMUT - vSERSYNT
d/dt(PEPc) = vPEPCK1 - vENOMUT
d/dt(OAAc) = - vGOT1 - vPEPCK1
d/dt(GLUTc) = - vGOT1 - vSERSYNT
d/dt(AKGc) = vSERSYNT + vGOT1
d/dt(ASPc) = vASPUP + vGOT1
```

We take mass action laws as rate functions, with the only dependence upon the internal metabolites (the external metabolites concentrations are equal to one):

```
vSEROUT = kSEROUT*SERc
vSERSYNT = kSERSYNT *s3PG*NADc*GLUTc
vENOMUT = kENOMUT*(PEPc - s3PG/KQENOMUT)
vPEPCK1 = kPEPCK1*OAAc*ATPc
vGOT1 = - kGOT1*(ASPc*AKGc - GLUTc*OAAc*KQGOT1)
vASPUP = kASPUP*ASP
```

The kinetic parameters are:

```
kSEROUT = 0.1
kSERSYNT = 0.1
kENOMUT = 0.1
KQENOMUT =1.056
kPEPCK1 = 0.1
kGOT1 = 0.03125
KQGOT1 = 3.2
kASPUP = 0.01
```

KQENOMUT = 1.056 and KQGOT1 = 3.2 are the equilibrium constants of the two reversible reactions of this network.
kSEROUT = 0.1, kSERSYNT = 0.1, kENOMUT = 0.1, kGOT1 = 0.03125 and kASPUP = 0.01 are the rate constants of the forward reaction.

The results are given on the Fig. 7 below which shows how the steady-state is reached. In this case the aspartate entry controls entirely the fluxes through the network (control coefficient = 1). All fluxes are equal to 0.01 or -0.01 (Fig. 7A). Figure 7B shows how the concentrations at steady-state are reached. The red curve represents the production of serine after a lag.
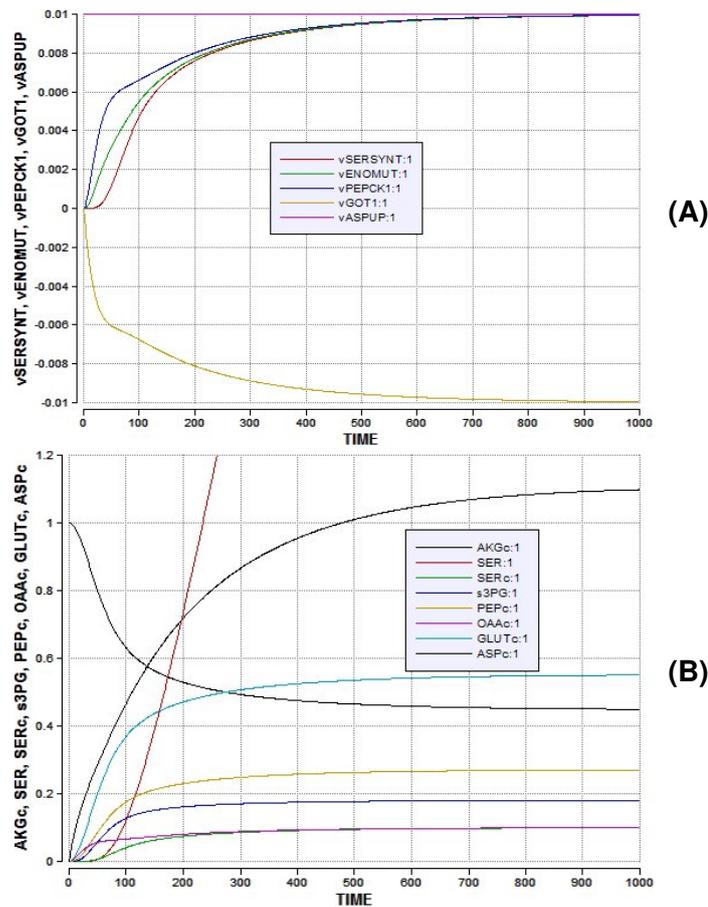
**Figure 7**:   Trajectories to steady state of (A) the fluxes and (B) the concentrations.

## 6   Conclusion

The great interest of a simple model of metabolism is that it can be approached by different methods: description of the minimal pathways (EFMs), FBA and FVA, dynamical behavior and control. We have shown that FBA and minimization of the fluxes give the simplest EFMs. The great number of EFMs (7018) despite the small size of the network (50 reactions and 27 internal metabolites) must be stressed. This helps to understand that it will be impossible to calculate all the EFMs of genome- scale models (GSM). In our case, not only are all the EFM describable but also they can be dynamically studied according to reasonable hypotheses on their kinetics.

It is obvious that a simple model cannot retain all the complexity of a GSM. Furthermore in the construction of the simple model some reactions may have been discarded without consequence in most of the conditions. However they could play an essential role in some peculiar conditions. For this reason, it is important to compare the behavior of the human-scale model (HSM) to the GSM as often as possible with the only method applicable to both models, i.e. the FBA

Our model HSM13 gives relevant results for central carbon metabolism. It will be completed by adding a simplified model of oxidative phosphorylation allowing to take into account the ATP usage and the redox balance of the cell and the mitochondria.

### *References*

[1] Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Price ND, Selkov E Sr, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JH, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BØ. *A community-driven global reconstruction of human metabolism.* Nat Biotechnol. 2013 May;31(5):419-25. doi: 10.1038/nbt.2488

[2] Bertrand Beauvoit, Sophie Colombié, Jean-Pierre Mazat, Christine Nazaret, Sabine Pérès. *Systematic Study of a Metabolic Network.* Proceedings of the Evry Spring School on "advances in Systems and Synthetic Biology". P. Amar, F. Képès and V. Norris ed. 2014.

[3] Schuster S, & Hilgetag C. (1994). *On elementary flux modes in biochemical reaction systems at steady-state.* Journal of Biological Systems, 2, 165-182.

[4] Schuster S, Fell DA, & Dandekar T. (2000). *A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.* Nature Biotechnology, 18, 326-332.

[5] Schilling CH, Letscher D and Palsson BO. *Theory for the systemic definition of metabolic pathway and their use in interpreting metabolic function from a pathway-oriented perspective.* J. theor. Biol.203 (2000) 229-248.

[6] `http://pinguin.biologie.uni-jena.de/bioinformatik/networks/`

[7] S. Pérès, F. Vallée, M. Beurton-Aimar, J.P. Mazat. *ACoM: A classification method for elementary flux modes based on motif finding*. BioSystems 103 (2011) 410–419

[8] Jeffrey D Orth, Ines Thiele & Bernhard Ø Palsson. *What is flux balance analysis?* Nature Biotechnology (2010) 28: 245-248.

[9] `http://f-a-m-e.fame-vu.vm.surfsara.nl/ajax/page1.php`

[10] Montal ED, Dewi R, Bhalla K, Ou L, Hwang BJ, Ropell AE, Gordon C, Liu WJ, DeBerardinis RJ, Sudderth J, Twaddel W, Boros LG, Shroyer KR, Duraisamy S, Drapkin R, Powers RS, Rohde JM, Boxer MB, Wong KK, Girnun GD. *PEPCK Coordinates the Regulation of Central Carbon Metabolism to Promote Cancer Cell Growth*. Mol Cell. (2015) 60:571-83

### *APPENDIX 1: Abbreviations*

**ANT**: ADP/ATP exchanger.
**AS**: ATP Synthase.
**ASPUP**: Uptake of aspartate.
**ATPASE**: ATPusage.
**CL**: Citrate Lyase.
**CS**: Citrate Synthase.
**ENOMUT**: Enolase + Phosphoglycerate Mutase.
**G1**: hexokinase + phosphoglucose isomerase.
**G2**: phosphofructokinase + aldolase + triose-phosphate isomerase.
**G3**: Glyceraldehyde-3P Dehydrogenase+ phosphoglycerate kinase.
**GG3**: triose phosphate isomerase + aldolase + fructose-1,6-biphosphatase.
**GG4** :phosphogluco isomerase + glucose-6-phosphatase.
**GLS1**:Glutaminase.
**GLNUP**: Uptake of Glutamine.
**GLUCUP**: Uptake of glucose.
**GLUD1**: Glutamate Dehydrogenase.
**GOT**:Glutamate Oxaloacetate Transaminases.
**K4**:Aconitase + Isocitrate dehydrogenase +.
**K567**: 2-oxoglutarate dehydrogenase + succinate thiokinase + succinate dehydrogenase + fumarase.
**L**: Leak of the membrane to protons.
**LACIO** : Input/Output of lactate.
**LDH**: Lactate Dehydrogenase.
**MAS**:Malate/Aspartate Shuttle.
**MDH**: Malate Dehdrogenase
**PDH**: Pyruvate Dehydrogenase.
**PEPCK**: PhosphoEnolPyruvate Carboxy Kinase.
**PK**: Pyruvate Kinase.
**PL1**: Synthesis of PhosphoLipids.
**PP1**: Oxidative part of PPP.
**PP2**: non-oxidative part of PPP.
**PPP**: Pentose Phosphate Pathway.
**PUR**: Purine Synthesis.
**PYC**: Pyruvate Carboxylase.
**RC**: Respiratory Chain.
**SEROUT**: Output of serine.
**SERSYNT**: Serine Synthesis= Dehydrogenase + Transaminase and Phosphatase.
**T2**(OGC): Glutamate/Aspartate exchanger.
**T4**(GLAST): Malate/2-oxoglutarate exchanger.
**T5**: Pi carrier.

ADVANCES IN SYSTEMS AND SYNTHETIC BIOLOGY                    55

## *APPENDIX 2: METATOOL entry file of RR13-3*

```
-ENZREV
AS ANT ENOMUT G3 GLUD1 GOT1 GOT2 K4 LACIO LDH MDH1
MDH2 ME1 ME2 NN PP2 T1 T2 T3 T4 T5 T7 T8 T9

-ENZIRREV
ASPUP ATPASE CL CS G1 G2 GG3 GG4 GLNUP GLS1 GLUCUP
GS1 K567 L PDH PEPCK1 PEPCK2 PK PL1 PP1 PUR PYC RC
SEROUT SERSYNT T6

-METINT
3PG ACoAc ACoAm AKGc AKGm ASPc ASPm CITc CITm G3P G6P
GLNc GLNm GLUCc GLUTc GLUTm LACc MALc MALm OAAc
OAAm PEPc PEPm PYRc PYRm R5P SERc

-METEXT
ADPc ADPm ASP ATPc ATPm BASES CO2 CoAc GLN GLUC Hc Hm
HCO3 LAC NADc NADHc NADm NADHm NADPc NADPHc NH3
Palmitate Pic Pim SER

-CAT
ANT : ATPm + ADPc = ATPc + ADPm .
AS : ADPm + Pim + 3 Hc = ATPm + 3 Hm .
ASPUP : ASP = ASPc .
ATPASE : ATPc = ADPc + Pic .
CL : CITc + ATPc + CoAc = ACoAc + OAAc + ADPc + Pic .
CS : ACoAm + OAAm = CITm .
ENOMUT : PEPc = 3PG .
G1 : GLUCc + ATPc = G6P + ADPc .
G2 : G6P + ATPc = 2 G3P + ADPc .
G3 : G3P + NADc + ADPc + Pic = 3PG + NADHc + ATPc .
GG3 : 2 G3P = G6P + Pic .
GG4 : G6P = GLUCc + Pic .
GLNUP : GLN = GLNc .
GLS1 : GLNm = GLUTm + NH3 .
GLUCUP : GLUC = GLUCc .
GLUD1 : GLUTm + NADm = AKGm + NADHm + NH3 .
GOT1 : GLUTc + OAAc = ASPc + AKGc .
GOT2 : GLUTm + OAAm = ASPm + AKGm .
GS1 : GLUTc + NH3 + ATPc = GLNc + ADPc + Pic .
K4 : CITm + NADm = AKGm + NADHm + CO2 .
K567 : AKGm + NADm + Pim + ADPm = MALm + NADHm
 + CO2 + ATPm .
L : Hc = Hm .
LACIO : LACc = LAC .
LDH : PYRc + NADHc = LACc + NADc .
```

```
MDH1 : MALc + NADc = OAAc + NADHc .
MDH2 : MALm + NADm = OAAm + NADHm .
ME1 : MALc + NADPc = PYRc + NADPHc + CO2 .
ME2 : MALm + NADm = PYRm + NADHm + CO2 .
NN : NADHc + NADm = NADHm + NADc .
PDH : PYRm + NADm = ACoAm + NADHm + CO2 .
PEPCK1 : OAAc + ATPc = PEPc + ADPc + CO2 .
PEPCK2 : OAAm + ATPm = PEPm + ADPm + CO2 .
PK : PEPc + ADPc = PYRc + ATPc .
PL1 : 8 ACoAc + 7 ATPc + 14 NADPHc + 7 HCO3 =
 Palmitate + 7
ADPc + 7 Pic + 14 NADPc + 8 CoAc + 7 CO2 .
PP1 : G6P + 2 NADPc = R5P + 2 NADPHc + 2 Hc + CO2 .
PP2 : 3 R5P = 2 G6P + G3P .
PUR : R5P + 2 GLNc + ASPc + 3 ATPc = BASES + 2 GLUTc
 + 3 ADPc + 3 Pic .
PYC : PYRm + HCO3 + ATPm = OAAm + Pim + ADPm .
RC : NADHm + 10 Hm = NADm + 10 Hc .
SEROUT : SERc = SER .
SERSYNT : 3PG + GLUTc + NADc = SERc + AKGc + NADHc .
T1 : CITm + MALc = CITc + MALm .
T2 : AKGc + MALm = AKGm + MALc .
T3 : MALm + Pic = MALc + Pim .
T4 : GLUTc + ASPm = GLUTm + ASPc .
T5 : Pic + Hc = Pim + Hm .
T6 : PYRc = PYRm .
T7 : PEPc = PEPm .
T8 : GLNc = GLNm .
T9 : GLUTc + Hc = GLUTm + Hm .
```

# Hybrid Gene Networks: a new Framework and a Software Environment

Emilien Cornillon[1], Jean-Paul Comet[1], Gilles Bernot[1], Gilles Enée[1]

[1] University Nice Sophia Antipolis, I3S laboratory - UMR CNRS 7271, 2000 route des lucioles, CS 40121, F-06903 Sophia Antipolis CEDEX, France

## *Abstract*

The modelling framework of René Thomas allows one to design abstract models of gene regulatory networks. In this formalism, time is also abstracted and dynamics are represented by the *succession* of discrete events. However, for numerous gene networks, the delay between two events is of first interest (circadian clock, cell growth, *etc*). In this chapter, we present a hybrid Thomas' formalism allowing us to take into account chronometrical information. We give the definition of the formalism and we present a first version of a user-friendly software platform named HyMBioNet. The main difficulty of an hybrid framework resides in the identification of the numerous parameters it involves. We illustrate our approach with an extremely simplified network of the mammalian circadian clock. For this example, we show how to determine accurate parameters. Finally, we show some simulations obtained *via* HyMBioNet.

## *1   Introduction*

Modelling a gene network consists in designing a virtual representation that provides a basis for the prediction of behaviours of interest. In a majority of cases, the key problem is the identification step which aims at determining the parameter values allowing a good representation of the known biological behaviours. A lot of network parameters are hard to identify and the larger the network, the more difficult the experimental determination of parameters. To overcome this difficulty, we gather information about the global behaviour of the biological complex systems, and we deduce some constraints on the parameters. Among information that can be experimentally measured, the elapsed time between two observed states is easy to evaluate and is often underused to model biological systems. In this article we propose a new hybrid formalism which gives the possibility to reflect the durations of regulations. Thus, measured elapsed times become a way to constrain the set of admissible models.

The differential equation frameworks [2, 13] rely on the hypothesis of continuity and homogeneousness of concentrations and provide very precise time

information. However, the domain of parameters is infinite. Thus, an automatic formal identification of parameters is very difficult especially when the differential system is not linear. On an other hand, the purely discrete formalisms [22, 14] make the reverse hypothesis: The concentration space is sliced into a small number of intervals, which reduces considerably the number of states and also reduces the number of values that the parameters can take. Thus the research space for parameters becomes finite and computer aided approaches based on formal logic help the modeller to find parameter values. For example, the Thomas' modelling framework [21] admits a completely automated identification step: The known behaviours are expressed in temporal logic formulas, all parameterisations are then checked by model-checking algorithm in order to select only the ones that are compatible with biological knowledge [6, 7]. Between these two classes of formalisms, some others exist [15, 1, 10] and finding a modelling frameworks that allow an automated reasoning and simplify the parameter identification step is an active research field.

The formalism presented in this chapter is an extension of the Thomas' formalism [21, 22]. Intuitively, whereas models of original Thomas' framework allow only discrete changes, the hybrid formalism presented here, in addition, mimics continuous evolutions of concentration inside each discrete state. Here, parameters can be thought as an evolution speed inside the discrete states, and the time to walk across the discrete state (at the considered speed) represents the delay mandatory inside the system to go through the discrete state.

We illustrate this formalism with the example of the mammalian circadian clock. It regulates a lot of important physiology mechanism [8]. Consequently, chronometrical time plays an important role in circadian cycle models. The circadian clock is well studied in mammals and it is often modelled with differential equations [17, 12, 3]. Each cell contains a clock which oscillates with its own period (approximately 24 hours) without stimuli. To oscillate in exactly 24 hours and synchronize with the others, cells receive *zeitgeber* inputs (synchronizers as light, temperature and food intake). The light is a very important zeitgeber and we will focus on it for the rest of this article.

The mammalian circadian clock is based on a gene regulatory network present in all cells. Here, we focus on the core of the molecular clock, composed of 4 elements: *Per* genes (*Per1, Per2, Per3*), *Cry* genes (*Cry1, Cry2*), *Bmal1* gene and *Clock* gene (or its homologue *Npas2*). Proteins of *Bmal1* and *Clock* form a complex CLOCK-BMAL1 which activates the transcription of *Per* and *Cry* genes. PER and CRY proteins are synthesized in the cytoplasm where they form complexes PER-CRY and accumulate. Once these complexes are phosphorylated, they move to the nucleus [24], where they are bound to CLOCK-BMAL1 and inhibit the activation of *Per* and *Cry* genes.

The presence of light induces an activation on *Per1* and *Per2* genes [19] by acetylation of the chromatin [5, 11] and an accumulation of PER-CRY complexes in the cytoplasm.

The chapter is organised as follows. In Section 2, we define our hybrid framework and the simplify circadian clock model on which we apply it. Section 3 presents a software platform to simulate our hybrid networks. Section 4 illustrates how one can build, in our running example, the constraints on the parameters. Section 5 shows some simulation results obtained with parameters satisfying the previous constraints. Finally, we discuss the adequacy of the approach.

## 2   A new hybrid formalism for gene networks

Our hybrid formalism is based on the René Thomas theory, enriched with multiplexes [16]. Section 2.1 gives the formal definitions and we explain the link between our parameters and the ones of Thomas. Section 2.2 presents the hybrid states and the notion of resources. Finally, Section 2.3 defines the hybrid state space of a network.

### 2.1   Hybrid gene networks

As shown in Figure 1, a gene network is visualized as a labelled directed graph (interaction graph) in which vertices are either variables (within circles) or multiplexes (within rectangles). Variables abstract genes and their products, and multiplexes contain propositional formulas that encode situations in which a group of variables (inputs of multiplexes, dashed arrows) influences the evolution of some variables (output of multiplexes, plain arrows). A multiplex can encode the formation of molecular complexes, phosphorylation by a protein, competition of entities for activation of a promoter, *etc*. To illustrate our formalism, we use a very simple abstraction of the mammalian circadian clock firstly defined in [9]. According to this abstraction, the main role of *Per* and *Cry* is to produce PER-CRY complexes and the effect of these complexes is to inhibit the *Per* and *Cry* genes. Thus, the model contains 2 variables, $g$ represents the genes *Per* and *Cry*, and $pc$ represent the complexes PER-CRY in the nucleus. The variable $g$ has a positive action on $pc$ (since the presence of $pc$ is a consequence of the activation of the clock genes) whereas $pc$ inhibits the variable $g$. During the day, light induces an accumulation of PER-CRY in the cytoplasm, preventing PER-CRY to enter the nucleus. This inhibits the inhibitory effect of $pc$ on $g$.

Figure 1 shows this model: the formula associated with the multiplex $m_{pc}$ is $(g \geqslant 1)$ which means that, when the expression level of $g$ reaches a certain value 1, it can help the activation of the variable $pc$. The other multiplex, $m_g$

is associated with the formula $\neg(pc \geqslant 1)$ which means on the contrary that $pc$ is an inhibitor of $g$ when it reaches the value 1. The variable $L$ (as Light) is a zeitgeber, an external variable that will be controlled during simulations. Consequently, it has no predecessor.
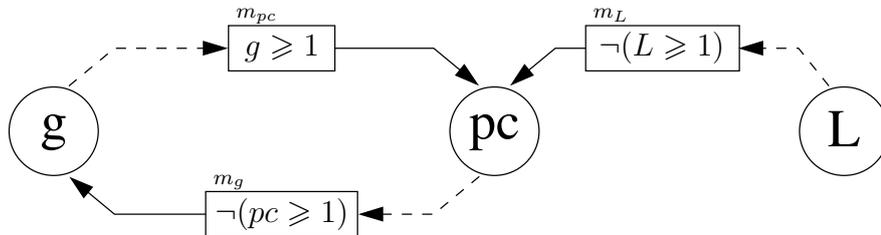


**Figure 1**: Interaction graph of the simplified gene regulatory network of the molecular circadian clock

**Notation 1.** *[sign function] The function* $sgn : \mathbb{R} \rightarrow \{-1, 0, 1\}$ *is defined by:*

$$sgn(x) = \begin{cases} \text{-1} & \text{if} \quad x < 0 \\ 0 & \text{if} \quad x = 0 \\ 1 & \text{if} \quad x > 0 \end{cases}$$

**Definition 1.** *[Hybrid gene regulatory network] A hybrid gene regulatory network (GRN for short) is a tuple* $R = (V, M, E, \mathcal{C})$ *where:*

- $V$ *is a set whose elements are called* variables *of the network. Each variable is associated with a boundary* $b_v \in \mathbb{N}^*$

- $M$ *is a set whose elements are called* multiplexes*. Each multiplex* $m \in M$ *is associated with a formula* $\varphi_m$ *belonging to the language* $\mathcal{L}$ *inductively defined by:*

  - *If* $v \in V$ *and* $n$ *is an integer such that* $1 \leqslant n \leqslant b_v$*, then* $v \geqslant n$ *is an atom of* $\mathcal{L}$

  - *If* $\varphi$ *and* $\psi$ *are two formulas of* $\mathcal{L}$*, then* $\neg\varphi$*,* $(\varphi \vee \psi)$*,* $(\varphi \wedge \psi)$ *and* $(\varphi \Rightarrow \psi)$ *also belong to* $\mathcal{L}$

- $E$ *is a set of edges of the form* $(m \rightarrow v) \in M \times V$*.*

- $\mathcal{C} = \{C_{v,\omega,n}\}$ *is a family of real numbers indexed by the tuple* $(v, \omega, n)$ *where* $v$*,* $\omega$ *and* $n$ *verify the three following conditions:*

  *1.* $v \in V$

2. $\omega$ is a subset of $R^-(v)$ where $R^-(v) = \{m \mid (m \rightarrow v) \in E\}$, that is $\omega$ is a set of predecessors of $v$.

3. $n$ is an integer such that $0 \leqslant n \leqslant b_v$

$C_{v,\omega,n}$ is called the celerity of $v$ for $\omega$ at the level $n$.

*Moreover, values of $C_{v,\omega,n}$ are constrained as in Figure 2. For each $v \in V$ and for each $\omega \subset R^-(v)$:*

- *Either all celerities $C_{v,\omega,n}$ with $0 \leqslant n \leqslant b_v$ have the same non null sign,*

- *or there exists $n_0$ such that $C_{v,\omega,n_0} = 0$ and for all $n$ such that $0 \leqslant n < n_0$, we have $sng(C_{v,\omega,n}) = 1$ and for all $n$ such that $n_0 < n \leqslant b_v$, we have $sng(C_{v,\omega,n}) = -1$.*

In the classical Thomas' formalism, a parameter $K_{v,\omega}$ is a value such that $0 \leqslant K_{v,\omega} \leqslant b_v$. They represent the value toward which the variable $v$ is attracted whereas celerities represent, in addition, the speed inside states. Celerities give more informations about the local dynamic of the model whereas the original Thomas' parameters only give global information. There exists a strong connexion between the original Thomas' parameters and our celerities (see Figure 2):

- If for all $n$, $C_{v,\omega,n}$ have a negative (resp. positive) sign, then $K_{v,\omega} = 0$ (resp. $K_{v,\omega} = b_v$)

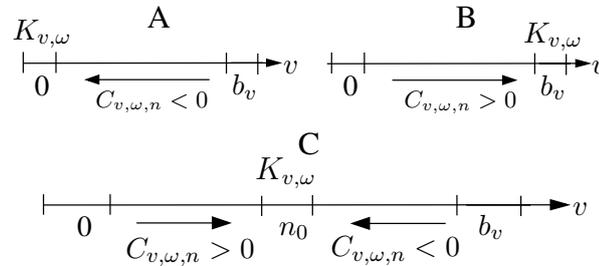- else $K_{v,\omega} = n_0$ (according to the previous definition)



**Figure 2**: Signs of celerities according to the value of Thomas' parameter

### Running example

Table 1 gives the network associated with Figure 1. This network has 8 qualitative states: 4 states $(g, pc) = (0, 1), (0, 1), (1, 0)$ and $(1, 1)$ in presence of light $(L = 1)$ and 4 other states in absence of light $(L = 0)$.

| $V:$ | $M:$ | $E:$ | $\mathcal{C}:$ | | |
|---|---|---|---|---|---|
| $g\,(b_g = 1)$ | $m_{pc}:\quad \varphi_{m_{pc}}$ | $(m_g \rightarrow g)$ | $C_{g,\{\},0}$ | $C_{pc,\{\},0}$ | $C_{pc,\{m_L\},0}$ |
| $pc\,(b_{pc} = 1)$ | $m_g:\quad \varphi_{m_g}$ | $(m_{pc} \rightarrow pc)$ | $C_{g,\{\},1}$ | $C_{pc,\{\},1}$ | $C_{pc,\{m_L\},0}$ |
| $L\,(b_L = 1)$ | $m_L:\quad \varphi_{m_L}$ | $(m_L \rightarrow pc)$ | $C_{g,\{m_g\},0}$ | $C_{pc,\{m_{pc}\},0}$ | $C_{pc,\{m_L,m_{pc}\},0}$ |
| | (see Figure 1) | | $C_{g,\{m_g\},1}$ | $C_{pc,\{m_{pc}\},1}$ | $C_{pc,\{m_L,m_{pc}\},1}$ |

**Table 1**: Simplified GRN of the molecular circadian clock, following Figure 1.

### 2.2 Hybrid state and resources

**Definition 2. [State of a GRN]** *Let $R = (V, M, E, \mathcal{C})$ be a GRN. A* hybrid state *of $R$ is a tuple $h = (\eta, \pi)$ where*

- *$\eta$ is a function from $V$ to $\mathbb{N}$ such that for all $v \in V$, $0 \leqslant \eta(v) \leqslant b_v$;*

  *$\eta$ is called the* discrete state *of $h$.*

- *$\pi$ is a function from $V$ to the interval $[0, 1]$ of real numbers,*

  *$\pi$ is called the* fractional part *of $h$*

*We denote $H$ the set of hybrid states of $R$. When there is no ambiguity, we often use $\eta$ and $\pi$ without explicitly mentioning $h$.*

The hybrid states combine two kinds of "states:" the discrete states (those of Thomas represented by $\eta$) and continuous positions inside the discrete states represented by $\pi$. As the width of a discrete state is 1, the domain of $\pi$ is the continuous interval $[0, 1]$.

**Definition 3. [Resources]** *Let $R = (V, M, E, \mathcal{C})$ be a GRN and let $v \in V$. The satisfaction relation $h \models \varphi$ (where $h = (\eta, \pi)$ is an hybrid state of $R$ and $\varphi$ a formula of $\mathcal{L}$) is inductively defined by :*

- *If $\varphi$ is the atom $v \geqslant n$ with $n \in [1, \cdots, b_v]$, then $h \models \varphi$ iff $\eta(v) \geqslant n$*

- *If $\varphi$ is of the form $\neg\psi$, then $h \models \varphi$ iff $h \nvDash \psi$*

- *If $\varphi$ is of the form $\psi_1 \vee \psi_2$, then $h \models \varphi$ iff $h \models \psi_1$ or $h \models \psi_2$*

- *If $\varphi$ is of the form $\psi_1 \wedge \psi_2$, then $h \models \varphi$ iff $h \models \neg(\neg\psi_1 \vee \neg\psi_2)$*

- *If $\varphi$ is of the form $\psi_1 \Rightarrow \psi_2$, then $h \models \varphi$ iff $h \models (\psi_2 \vee \neg\psi_1)$.*

*The set of resources of a variable $v$ for a state $h$ is defined by: $\rho(h, v) = \{m \in R^-(v) \mid h \models \varphi_m\}$, that is, the multiplexes predecessors of $v$ whose formula is satisfied.*

The notion of resources is important because it defines which celerity has to be considered in each discrete state: For our example, the celerity $C_{pc,\{m_{pc}\},1}$ (see Table 2) is used for the evolution of $pc$ in the discrete states where $\eta(pc) = 1$ and the set of resources of $pc$ is $\{m_{pc}\}$, that is where $\eta(g) \geqslant 1$ and $\eta(L) \geqslant 1$.

**Remark 1.** *The values of celerities are not proportional to the actual reaction speed in biology because we consider all discrete states with the same size (from 0 to 1) whereas, from a concentration point of view, the intervals of concentration can be of very different sizes.*

### *2.3  Hybrid trajectories*

We now define the trajectories of the model, both inside a discrete state and at the crossing of thresholds (between two discrete states). We firstly define the touch delay, measuring the time necessary to touch the boundary of the current discrete interval.

**Notation 2.** *[Touch delay] Let $R = (V, M, E, \mathcal{C})$ be a GRN, $v$ be a variable of $V$ and $h = (\eta, \pi)$ be a hybrid state, we note $\delta_h(v)$ the touch delay of $v$ in $h$ for reaching the border of the discrete state. More precisely, $\delta_h$ is the function from $V$ to $\mathbb{R}^+ \cup \{+\infty\}$ defined by:*

- *If $C_{v,\rho(v,\eta),\eta(v)} = 0$ then $\delta_h(v) = +\infty$*

- *If $C_{v,\rho(v,\eta),\eta(v)} > 0$ then $\delta_h(v) = \frac{1-\pi(v)}{C_{v,\rho(v,\eta),\eta(v)}}$*

- *If $C_{v,\rho(v,\eta),\eta(v)} < 0$ then $\delta_h(v) = \frac{\pi(v)}{|C_{v,\rho(v,\eta),\eta(v)}|}$*

If the celerity is null, the variable cannot evolve, thus the touch delay is infinite. Else, the touch delay depends on the associated celerity and the value $\pi(v)$.
$\delta_h(v) = 0$ means that the trajectory arrived on a border of the discrete state.

Let us notice that $v$ may never reach its border: It is the case when another variable reaches its border before $v$ (that is with a shorter $\delta_h$).

**Definition 4.** *[Black wall and boundary] Let $R = (V, M, E, C)$ be a GRN, let $v \in V$ be a variable, let $h = (\eta, \pi)$ and $h' = (\eta', \pi')$ be two hybrid states such that $\eta$ and $\eta'$ are neighbour states w.r.t. $v$,
that is $\eta'(v) = \eta(v) + sgn(C_{v,\rho(v,\eta),\eta(v)})$ and $\eta'(u) = \eta(u)$ for all $u \neq v$.*

1. *$v$ is said on a* black wall *if the two following conditions are satisfied:*

   - *$\delta_h(v) = 0$*
   - *$sgn(C_{v,\rho(v,\eta),\eta(v)}) \times sgn(C_{v,\rho(v,\eta'),\eta'(v)}) = -1$*

2. $v$ is said on a boundary *if the two following conditions are satisfied:*

- $\delta_h(v) = 0$
- *At least one of the following conditions holds:*
  - $C_{v,\rho(v,\eta),\eta(v)} < 0$ *and* $\eta(v) = 0$
  - $C_{v,\rho(v,\eta),\eta(v)} > 0$ *and* $\eta(v) = b_v$.

*We note* $slide(h)$ *the set of variables that are either on a black wall or on a boundary.*

We can remark that for each hybrid state, the continuous component takes its values in the interval $[0,1]$, leading to a sort of a "double threshold" between two neighbour discrete states, as $h(v) = (n,1)$ differs from $h(v) = (n+1,0)$. The behaviour of the system will use this "double threshold" for sliding states.

For example, it becomes possible to slide on a "black wall" (Definition 4.1). The trajectories from one of these discrete states do not cross the threshold (because on the opposite side, the celerities make impossible an entering trajectory). See Figure 3A.
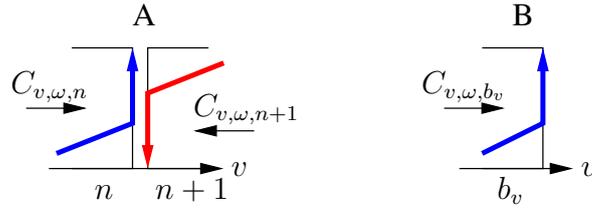


**Figure 3**: (A) Example of a black wall. (B) Example of a boundary

Also, in a discrete state such that $\eta(v) = b_v$ (resp. $\eta(v) = 0$) where $C_{v,...} > 0$ (resp. $C_{v,...} < 0$), the variable $v$ cannot cross the boundary (there is no state on the other side). Then, other variables can allow the trajectory to slide on the boundary. See Figure 3B.

**Definition 5.** *[Knocking variables] Let* $R = (V, M, E, C)$ *be a GRN and* $h = (\eta, \pi)$ *be a hybrid state, the set of* knocking variables *is defined by :*

$first(h) = \{v \in V \smallsetminus slide(h) \mid \delta_h(v) < +\infty$ *and* $\forall u \in V \smallsetminus slide(h)$, $\delta_h(u) \geqslant \delta_h(v)\}$

The set $first(h)$ represents the set of variables whose qualitative value can change first. If the variable is on a sliding wall, it cannot evolve as long as other variables do not change. Similarly, if the celerity of a variable $v$ in the current state is null, its qualitative value cannot change because of its infinite delay.

**Definition 6.** *[Hybrid state space] Let $R = (V, M, E, C)$ be a GRN, we note $\mathcal{R} = (H, T)$* the hybrid state space *of $R$ where $H$ is the set of hybrid states and $T$ is the set of* transitions*: There exists a transition from $(\eta, \pi)$ to $(\eta', \pi')$ iff there exists a variable $v \in first(h)$ such that*

1. *Either $\delta_h(v) \neq 0$ and*

   (i) $\eta' = \eta$

   (ii) $\pi'(v) = \frac{1 + sgn(C_{v,\rho(v,\eta),\eta(v)})}{2}$ *for all $v' \in first(h)$*
   *(i.e. 0 if $C_{v,\rho(v,\eta),\eta(v)} < 0$ and 1 if $C_{v,\rho(v,\eta),\eta(v)} > 0$).*

   (iii) *For all variables $u \neq v$, if $u \notin slide(h)$ then $\pi'(u) = \pi(u) + \delta_h(v) \times C_{u,\rho(u,\eta),\eta(u)}$, else $\pi'(u) = \pi(u)$.*

2. *or $\delta_h(v) = 0$ and*

   (i) $\eta'(v) = \eta(v) + sgn(C_{u,\rho(v,\eta),\eta(v)})$

   (ii) $\pi'(v) = \frac{1 - sgn(C_{v,\rho(v,\eta),\eta(v)})}{2}$ *(i.e. 0 if $C_{v,\rho(v,\eta),\eta(v)} > 0$ and 1 if $C_{v,\rho(v,\eta),\eta(v)} < 0$).*

   (iii) *For all variables $u \neq v$, $\eta'(u) = \eta(u)$ and $\pi'(u) = \pi(u)$*

There are two different kinds of transitions:

1. Inside a discrete state: The idea is to determine a next hybrid state which could give rise to a qualitative change. Thus, from a hybrid state, one has to determine the variables which first reaches the border of the discrete state. The resulting hybrid state is then the one where the fractional part of $v$ is equal to 0 or 1 (see $1.ii$), and other variables are changed accordingly. Inside a given discrete state, all trajectories of the hypercube are parallel ones because all hybrid states have the same celerities.

2. Between two discrete states: If the system cannot evolve anymore within the current discrete state, the trajectory goes through a border: The discrete part is computed according to the sign of the celerity (see $2.i$) and the fractional part is either 1 or 0 according to the sign of celerities (see $2.ii$). Other variables are not changed.

This semantics has been inspired by the trajectories produced by the usual stepwise linear ordinary differential equation systems [20]. In the neighbourhood of a black wall, the trajectories on one side of the threshold are issued from the differential system of *this* side. In other words, even if the trajectory remains very close to the threshold: the evolution depends on the domain from which the trajectory comes. This is why in our semantics, all slidings on a black wall, take place inside the discrete state from which the trajectory comes.

### *Running example*

To illustrate the transitions, let us consider the hybrid state $d$ of Figure 12B. $d$ is of the form $d = ((1,0),(0,d_y))$: the discrete state is $(\eta(g) = 1, \eta(pc) = 0)$ and the fractional part is $(\pi(g) = 0, \pi(pc) = d_y)$ where $0 < d_y < 1$.

In this discrete state, both celerities are positive. In the figure, we see that $\delta_d(g) < \delta_d(pc)$. So $first(d) = \{g\}$. Thus, the trajectory reaches the boundary of $g$ on which $g$ is able to slide (hybrid state $d_1$ in the figure). The rest of the trajectory follows the same reasoning according to Definition 6: It slides along the boundary according to the celerity of $pc$ and reaches state $f$. The trajectory crosses the $pc$ threshold and enters into the next discrete state. In the next discrete state, we have $\delta_a(g) > \delta_a(pc)$, thus the boundary of $pc$ is first reached. The trajectory slides on the boundary of $pc$ according to the celerity of $g$ to reach the threshold of $g$ and finally reach the hybrid state $b$.

## 3   The HyMBioNet software

The main difficulty when modelling GRN lies in the parameter identification. Model checking has been widely used in purely discrete frameworks to select parameters allowing the model to fulfil all the known biological behaviours [7]. However, in hybrid formalisms, this method leads to very complex procedures [23]. In practice, we currently use CTL to express known biological behaviours and we use model checking *via* SMBioNet [7] to select parameters of the purely discrete Thomas' model. We then automatically deduce the sign of celerities of our hybrid model (see Section 2.1). To go further, we choose to perform simulations within the hybrid model. In this aim, we developed a user-friendly software platform to generate and to simulate a network in order to confront the model with biological expertise. The HyMBioNet platform consists in two tools: The first one allows the construction of a GRN through a web interface and generates the code of a simulator in the NetLogo language [25]. This simulator can then be used *ad libitum*. We describe these two tools in this section.

### 3.1   The HyMBioNet web interface

In order to build a GRN, we first have to list the genes and we also allow the use of *zeitgeber*. A zeitgeber is an external variable that influences upon gene(s); such a variable behaves independently of the rest of the network. For example, the night and day alternance is a zeitgeber for a circadian clock gene network.

### Adding genes and zeitgebers

Figure 4 shows the main web application. The user is repeatedly invited to fill a gene name, to indicate the associated boundary and then to press the "Add Gene" button. Zeitgebers can be added thanks to the same mechanism in the "Add new zeitgeber" filling zone.



**Figure 4**: HyMBioNet web interface

### Adding multiplexes

When the genes and the zeitgebers have been given, the user is invited to describe the multiplexes. Figure 5 gives the description step by step of how to create the multiplex $Ml$, i.e., the multiplex that explains the influence of the "Light" zeitgeber on $pc$ in our model. First, the name of the multiplex is filled in the "Rule name" field (Figure 5A); note that this field appears when there is at least one gene in the network. Then the user is invited to write the formula associated with this multiplex thanks to a user friendly interface that prevents errors by guiding the user along the syntactic tree of the formula (5-B). Here the formula of the multiplex $Ml$ is $\neg(L \geq 1)$ and it has the gene "$pc$" as target. In step 5-C, the user has chosen $L$ as the variable that controls the multiplex $Ml$. In step 5-D, the user selects the comparator "$\geqslant$". Next, (s)he selects the value of $L$ used for the comparison: only the possible values corresponding to $L$ are shown (step 5-E). As the formula of this multiplex contains a negation, the "Not ?" checkbox has to be checked and then the "Not or not Not ?"

button has to be pressed (see 5-F). To finish multiplex description, the target of the multiplex is checked (here the variable $pc$) and the button "Update Rule Targets" is pressed: the final multiplex is displayed in Figure 5G.
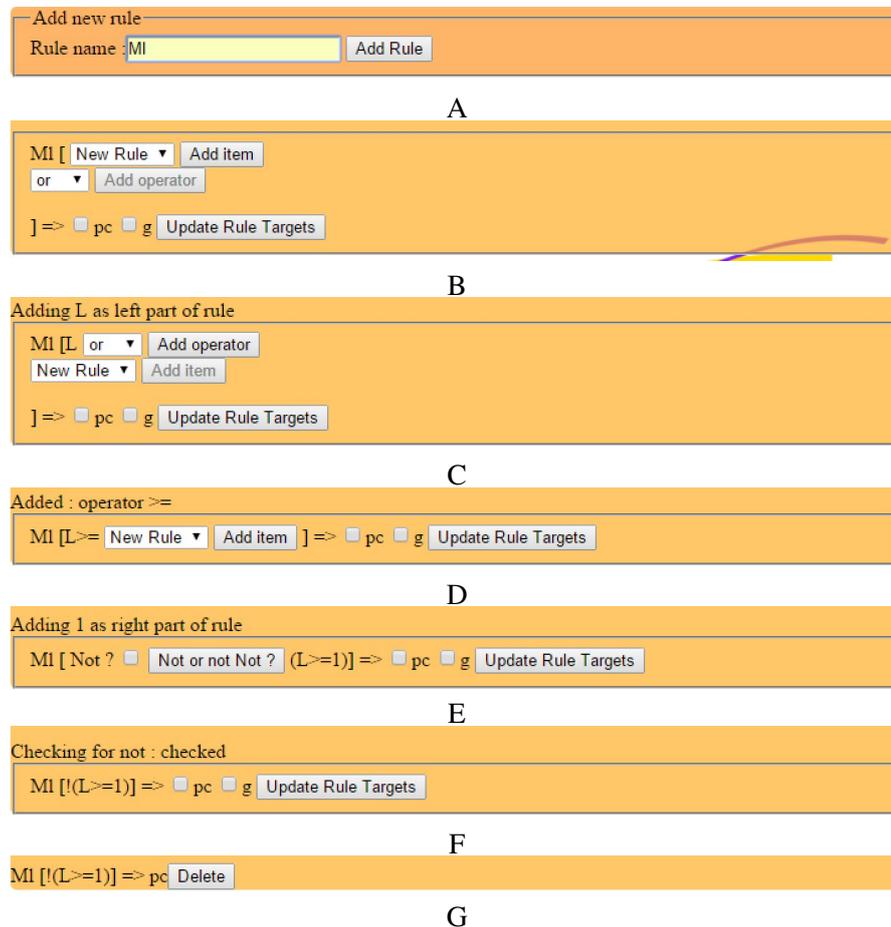


**Figure 5**: Interface for creating a multiplex

### Automatically added Celerities

Whereas genes, zeitgebers and multiplexes are acquainted by the user, celerities described in Section 4 are automatically listed by the application. Figure 6 shows the "display text" version of the simple mammalian circadian model. The number of celerities associated with a gene depends on the number of multiplexes that regulate this gene and also on the qualitative levels of this

gene (see Definition 1). There is a celerity for each subset of regulations (if $g$ is regulated by $n$ multiplexes, there are $2^n$ possible subsets) and for each qualitative level. All celerities are produced regarding multiplex targets that are described in the "REG" part.



**Figure 6**: Simple model of the mammalian circadian clock

For example, $pc$ is regulated by multiplexes $Mpc$ and $Ml$. There are 4 possible subsets of regulators: $\varnothing$, $\{Mpc\}$, $\{Ml\}$ and $\{Mpc, Ml\}$. This gives rise to the celerities C_pc[], C_pc[Mpc], C_pc[Ml] and C_pc[Mpc,Ml], but $pc$ can take two distinct qualitative levels, so we get the 8 celerities of Figure 6 (see green zone).

### Features

We added the possibility to save and upload a model. This is very useful to test different options. As a second enhancement, we offer to delete useless celerities (that are never used because of the unsatifiability of the associated multiplex formulas) in order to produce a lighter simulation.

### Export your simulation

Lastly, when modelling is complete, one can export the corresponding NetLogo simulator with the "Generate NetLogo Simulation" button.

### 3.2 NetLogo simulation

The NetLogo simulation is shown in Figure 7. The black board is where the GRN is displayed and animated.

On the left side, there are three buttons. One to "Setup" a simulation, the second to make a simulation "Step by step" and the third to make a "Loop" simulation. While running, plots are automatically updated with data, for each

variable: The level of each variable is represented by the algebraic sum of the current qualitative level and the current position inside this discrete state. Plots can be exported in CSV format thanks to export buttons.

On the right side of the central black board, it is possible to change the number of simulation steps per hour: if this slider is set to 1, there is a unique simulation step for each hour, and if it is set to 60, each step of simulation corresponds to 1 minute between two clock ticks. Under this slider, three other sliders allow the user to set the initial gene / zeitgeber values. The current value is shown on the left of the slider.
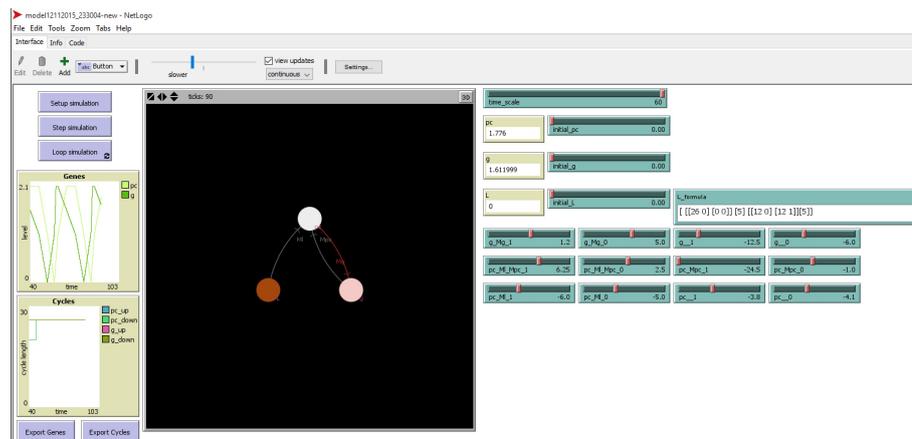


**Figure 7**: Netlogo Simulation

### *Zeitgeber language description*

As zeitgebers are specific variables that are able to influence the network but are not influenced by it, we propose a small language to describe the time evolution of zeitgebers.

The Language is the following:

- $formula ::== [daylist]$

- $daylist ::== day \mid day\ daylist$

- $day ::== [pair\_value\_list][how\_many\_times]$

- $pair\_value\_list ::== pair\_value \mid pair\_value\ pair\_value\_list$

- $pair\_value ::== [hours\ value]$

where $hours$ is the number of hours the zeitgeber will stay in the specified discrete $value$ and $how\_many\_times$ is the number of times the previous described cycle is repeated. For example, in Figure 7, the expression `[[[12 0] [12 1]][1]]` indicates that the zeitgeber will be at the discrete level 0 for 12 hours, then it will change to discrete value 1 for the next 12 hours and finally that cycle is not repeated. When the cycle is finished, i.e., the day has passed here, the system will restart that cycle from the beginning.

### *Fixing celerities*

The simulator manages delays instead of celerities (delays are easier to handle). Because the length of each qualitative interval is one, the mandatory delay to cross an interval is equal to $\frac{1}{celerity}$. Thanks to sliders varying in $[-25; 25]$ with a $0.1$ increment, one can fix each delay of the described model. Delays are named in the same manner than celerities. The settings corresponding to celerities of Table 2 are shown in Figure 8.



| g_Mg_1 | 1.2 | g_Mg_0 | 5.0 | g__1 | -12.5 | g__0 | -6.0 |
| pc_Ml_Mpc_1 | 6.25 | pc_Ml_Mpc_0 | 2.5 | pc_Mpc_1 | -24.5 | pc_Mpc_0 | -1.0 |
| pc_Ml_1 | -6.0 | pc_Ml_0 | -5.0 | pc__1 | -3.8 | pc__0 | -4.1 |
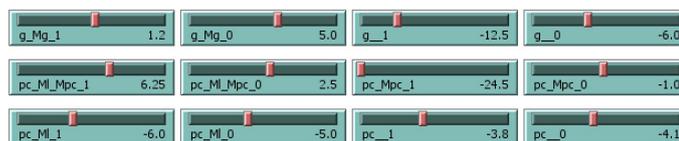
**Figure 8**: Chosen delays

### *Running example*

After delays are fixed, the simulation must be "setup" in order to display genes, zeitgeber and multiplexes (see Figure 9). Simulation is launched using the "Loop simulation" button while a step by "step simulation" button is available to slowly observe transitions.

Figure 10A shows that the length of the mammalian circadian cycle is, as expected, about 26 hours in constant dark. In Figure 10B, the cycle is 24 hours when exposed to a 12h/12h alternation of dark and light. In that case, $g$ gene does not reach the 0 sliding wall.

**Figure 9**: Main simulator view after setup



**Figure 10**: Circadian cycle simulation

To produce those plots, the formula controlling the time evolution of the zeitgeber is: `[[[26 0][0 0][5][[12 0][12 1]][5]]`. Five 12h/12h alternations of dark/light follow five cycles of 26 hours in constant dark.

### *Drawback*

If a network has a large number of parameters, sliders will not all fit into the screen space and will only be accessible *via* scroll bars.

## 4  *Identification of celerities*

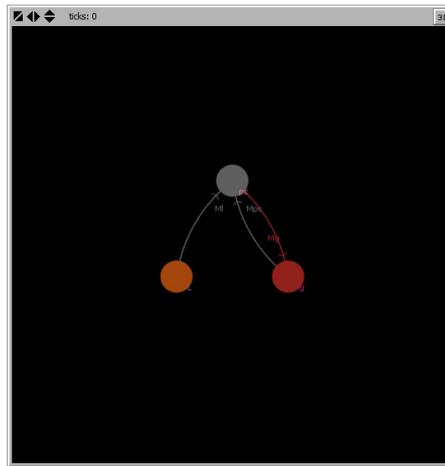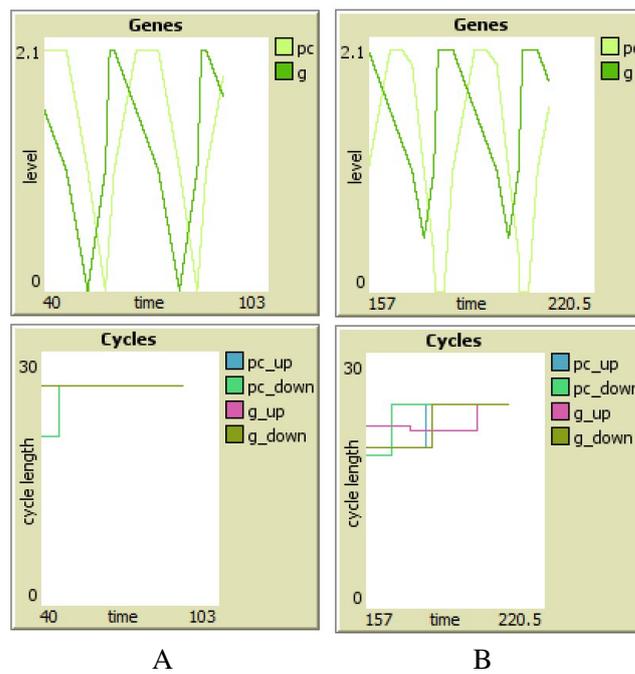As usual, a key point of the modelling process is the parameters identification. This identification has to be done taking into account the known behaviours of the biological system.

On our example, the crucial and the simplest property that the model has to exhibit is the following: the circadian clock has sustained oscillation, even if no zeitgeber controls the oscillation. Of course we can make a lot of simulations using HyMBioNet in order to get an intuition about the celerities allowing the model to behave as expected. However, we would like to provide a full analysis in order to build constraints on celerities making possible a sustained oscillation in constant dark (in the plane $L = 0$). Following the preliminary model of the circadian clock [9], we deduce the sign of all celerities and then the direction of the rotation of the cycle: in counterclockwise direction $((0,0) \to (1,0) \to (1,1) \to (0,1) \to (0,0))$, see Figure 12.

In this section, we firstly prove that in order to exhibit a unique attractive limit cycle, the limit cycle has to slide on a boundary. In Section 4.2, we use the previous property to build a relationship between the different slopes of the cyclic trajectory. Finally, Section 4.3 uses the known period in order to constrain celerities.

### 4.1  *Using 2D remarkable properties*

Let us note respectively $slope_{00}, slope_{01}, slope_{10}, slope_{11} \in I\!\!R_+^*$ the "slope" of trajectories inside each discrete state $(0,0), (0,1), (1,0), (1,1)$.

- in $(0,0)$, $slope_{00} = \dfrac{-C_{pc,\{m_{pcl}\},0}}{C_{g,\{m_g\},0}} > 0$

- in $(0,1)$, $slope_{01} = \dfrac{C_{pc,\{m_{pcl}\},1}}{C_{g,\varnothing,0}} > 0$

- in $(1,0)$, $slope_{10} = \dfrac{C_{pc,\{m_{pcl},m_{pcg}\},0}}{C_{g,\{m_g\},1}} > 0$

- in $(1,1)$, $slope_{11} = \dfrac{C_{pc,\{m_{pcl},m_{pcg}\},1}}{-C_{g,\varnothing,1}} > 0$

Let us remark that for geometrical facility, we take the absolute value of the mathematical slopes.

**Proposition 1.** *Let $R = (V, M, E, \mathcal{C})$ be a GRN where $|V| = 2$ and the boundaries of both variables are equal to 1. There exists a unique attractive limit cycle if and only if there exists a cyclic trajectory that slides on at least one external boundary.*

*Proof.* Necessary condition: If there is no place where the cyclic trajectory slides, then because all slopes are constant in each of the four discrete states, there would be an infinity of different cyclic parallel trajectories (Figure 11A):
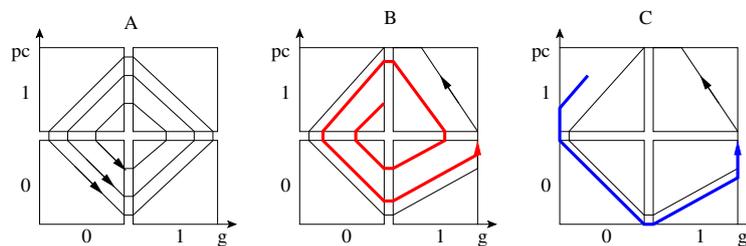


**Figure 11**: (A) State graph with an infinity of cyclic trajectories. (B) State graph with trajectory inside cycle. (C) State graph with external trajectory

Thus, there is at least one sliding.
Sufficient condition: If there exists a cyclic trajectory that slides on an external boundary then it is unique and attractive because:

- For any hybrid state inside the cycle, its trajectory after exactly one rotation (Figure 11B) will lead to a state that is strictly closer to the surrounding cycle (divergent red "spiral"). The sequence of such points after $n$ rotations will cross the limit cycle.

- Similarly, for any hybrid state outside the cycle, its (blue) trajectory will necessary stay outside the cycle because, in 2D, trajectories never cross each other. Consequently, its trajectory will necessary slide where the limit cycle slides, and join the limit cycle after at most one rotation (Figure 11C).

$\square$

### 4.2   Constraints on the slopes of the cyclic trajectory

Here we would like to build some constraints on celerities in order to facilitate the choice of "valuable" celerities. The high synthesis rate of PER and CRY proteins leads to an accumulation of PER-CRY complexes in the cytoplasm.

This accumulation induces a fast crossing of complexes into the nucleus, so, a saturation of the nucleus by PER-CRY complex [9]. According to these informations, we impose 2 slidings in our model.

- The cyclic trajectory slides in the discrete state $(1, 0)$ for the maximal rate of the transcription of clock genes and

- it also slides in $(1, 1)$ for the saturation of PER-CRY complex in the nucleus.

Figure 12B represents the expected trajectory in constant dark conditions.
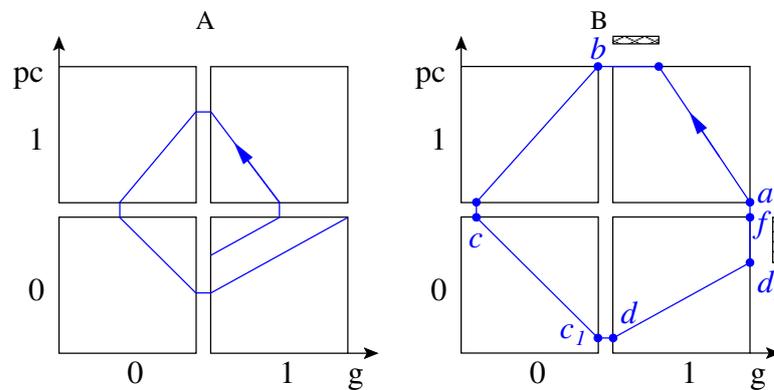


**Figure 12**: Trajectories of the simplified circadian clock model in constant dark. (A) A trajectory with an initial state inside the cycle. (B) The limit cycle.

As we have fixed the slidings on states $(1, 0)$ and $(1, 1)$, we know that the first hybrid state of the cyclic trajectory in $(1, 1)$ has $(1, 0)$ for fractional coordinates(see Figure 12B). To have a sliding in the state $(1, 1)$, $pc$ must touch its boundary before $g$.

1. Constraint for the discrete state $(1, 1)$: The trajectory must slide on the threshold of $pc$, so $slope_{11} > 1$.

2. Constraints on $(0, 1)$. In this discrete state, the trajectory begins at the hybrid state $b$. As the trajectory does not reach the boundary of $g$, $pc$ must reach its threshold before $g$. Thus, we have $slope_{01} \geqslant 1$.

3. Constraints on $(0, 0)$. The fractional coordinates of the hybrid state $c$ are $(-slope_{01}, 1)$ and those of the state $c_1$ are $(0, -slope_{01} \times slope_{00})$. This leads to the constraints $slope_{00} \times slope_{01} \leqslant 1$.

4. Constraints on $(1,0)$. The fractional coordinates of the hybrid state $d$ are $(0, -slope_{00} \times slope_{01})$. As this discrete state contains a sliding on the boundary of $g$, the trajectory must reach it before the threshold of $pc$. This leads to the constraints $slope_{00} \times slope_{01} > slope_{10}$.

Among the constraints, we keep the first constraint $slope_{11} > 1$ for the sliding inside the discrete state $(1,1)$ and the last constraint $slope_{00} \times slope_{01} > slope_{10}$ for the sliding inside the discrete state $(1,0)$ depending on the previous states $(0,0)$ and $(0,1)$.

### 4.3 Using the length of the period of the circadian clock in constant dark

**Notation 3.** *Let us denote by $\delta_{00}^g, \delta_{01}^{pc}, \delta_{10}^{pc}, \delta_{11}^g \in I\!\!R_+^*$ the times to cross each states $(0,0), (0,1), (1,0), (1,1)$ respectively.*

These crossing delays depend on the only variable that can change its qualitative value and its celerity. For example, in the state $(0,0)$, only the variable $pc$ can evolve because $g$ goes toward its boundary and slides on it. Thus, the only delay to be taken into account is the delay of $pc$.

Without the zeitgeber, the clock have a period between 22 and 25 hours depending of the tissue [26]. Here, we voluntarily chose a longer period of 26 hours to an easier observation of the effect of zeitgeber (the light) on the synchronisation. Thus we impose: $\delta_{00}^g + \delta_{10}^{pc} + \delta_{11}^g + \delta_{01}^{pc} = 26$.

- For the discrete state $(1,1)$, $\delta_{11}^g = \frac{1}{C_{g,\rho(a,g),1}}$ because the crossing distance of $g$ is equal to 1.

- For the discrete state $(0,1)$, for the same reasons, we have $\delta_{01}^{pc} = \frac{1}{C_{pc,\rho(c,pc),1}}$.

- In $(0,0)$, the crossing distance is equals to $slope_{01}$. Thus, we have $\delta_{00}^g = \frac{slope_{01}}{C_{g,\rho(d,g),0}}$.

- In $(1,0)$, the crossing distance of $pc$ equals to $slope_{00} \times slope_{01}$, so we have $\delta_{10}^{pc} = \frac{slope_{00} \times slope_{01}}{C_{pc,\rho(e,pc),0}}$.

At the end, to have a cycle in constant dark, celerities must satisfy the following constraint :

- $slope_{10} > slope_{00} \times slope_{01}$

- $slope_{11} > 1$

- $\frac{1}{C_{g,\rho(a,g),1}} + \frac{1}{C_{pc,\rho(c,pc),1}} + \frac{slope_{01}}{C_{g,\rho(d,g),0}} + \frac{slope_{00} \times slope_{01}}{C_{pc,\rho(e,pc),0}} = 26$

The Table 2 contains the chosen celerities to make the simulations of Section 5.

| $g$ | | | $pc$ | | |
|---|---|---|---|---|---|
| $C_{g,\{\},0}$ | $=$ | $-0.167$ | $C_{pc,\{\},0}$ | $=$ | $-0.244$ |
| $C_{g,\{\},1}$ | $=$ | $-0.08$ | $C_{pc,\{\},1}$ | $=$ | $-0.262$ |
| $C_{g,\{m_g\},0}$ | $=$ | $0.2$ | $C_{pc,\{m_{pc}\},0}$ | $=$ | $-1.0$ |
| $C_{g,\{m_g\},1}$ | $=$ | $0.8$ | $C_{pc,\{m_{pc}\},1}$ | $=$ | $-0.041$ |
| | | | $C_{pc,\{m_L\},0}$ | $=$ | $-0.2$ |
| | | | $C_{pc,\{m_L\},1}$ | $=$ | $-0.167$ |
| | | | $C_{pc,\{m_L,m_{pc}\},0}$ | $=$ | $0.4$ |
| | | | $C_{pc,\{m_L,m_{pc}\},1}$ | $=$ | $0.16$ |

**Table 2**: Chosen parameters values according to constrains determined in Section 4.

## 5  Some results

In the sequel, the couple $(\eta(v), \pi(v))$ where $(\eta, \pi)$ is a hybrid state, describes the qualitative level and the fractional part of variable $v$, and such a couple is represented by the algebric sum $\eta(v) + \pi(v)$. Thus the real number $1.5$ represents the couple $(1, 0.5)$. However the real number $2.0$ represents both couples $(1, 1.0)$ and $(2, 0.0)$. In other words, two points on both sides of a threshold are merged.

In Figures 13, 14 and 15, red (resp. blue) curves represent the evolution of $pc$ (resp. $g$), whereas the grey zones correspond to the dark phases ($L = 0$). The dashed line represents the threshold of each variable. Used parameters are those of Table 2.

### 5.1  Light as zeitgeber: constant dark vs. equinox

The Figure 13A represents the simulation in constant dark conditions. The sustained oscillations have a 26-hours period. Each sliding imposed in the previous section is observable in the figure: It is represented by the horizontal segments where $g = 2$ (or where $pc = 2$). Figure 13B is a zoom on a single cycle of Figure 13A. Figure 13C, shows the behaviour of the model with the adding of the dark/light oscillation with 12 hours of light and 12 hours of dark (equinox). During the light phase, the decrease is stronger than during the dark phase (B).

This disruption allows the model to reach, before the end of the cycle (between 23 and 24 hours), a state where neither $g$ (blue) nor $pc$ (red) can move anymore. We call it the "waiting state" because this state remains stable until the next dark phase starts and allows the system to synchronise to the stable period imposed by the zeitgeber.
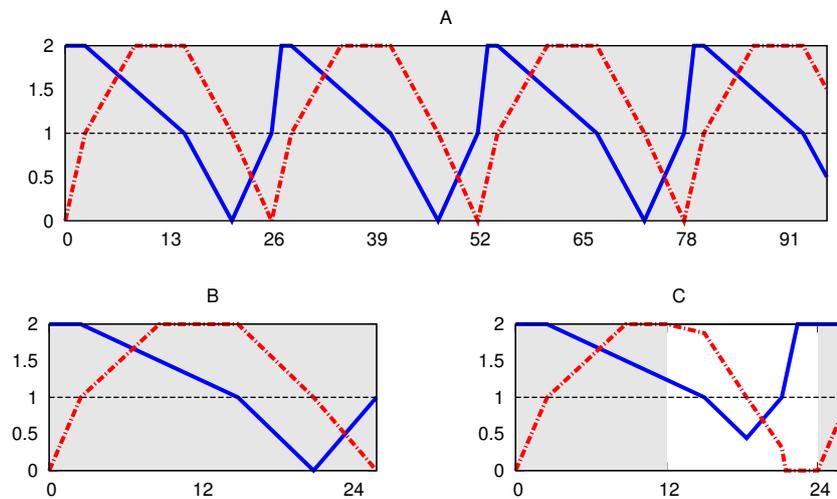
**Figure 13**: (A) Full simulation in constant dark (B) Zoom on a single cycle in constant dark (C) Simulation of the model submitted to a 12h/12h oscillations of dark and light.

### 5.2   *Circadian clock and seasons*

Figure 14 shows the behaviour of the model submitted to different seasonal conditions. In conditions $A$ and $B$, corresponding respectively to summer and winter, in temperate zone, the amplitude of the curves of $g$ and $pc$ changes slightly but the system is able to maintain a periodic behaviour of 24 hours. The waiting state plays the role of a "buffer" to make the system robust to the dark/light alternation disruption.

- In extreme winter condition ($D$), the waiting state disappears and the end of the cycle happens after 24 hours. Thus, the system tends toward a 26 hours period as in constant dark period: The light phase is not long enough to synchronize the circadian clock with a 24-hours period.

- In extreme summer condition ($C$), the amplitude of the curves decrease as in $B$, but this time, $g$ does not cross its threshold anymore. The system oscillates between the discrete states $(1, 0)$ and $(1, 1)$ (for $(g, pc)$) thus the variable $g$ is not repressed anymore.

Schmal *et al.* [18] present a theoretical study of seasonality in which they focus on entrainment of the circadian clock under different conditions. In particular, they present a "Arnold's onion" which show the entrainment zone of the circadian clock according to the period and the photoperiod (percentage of dark in 24 hours).
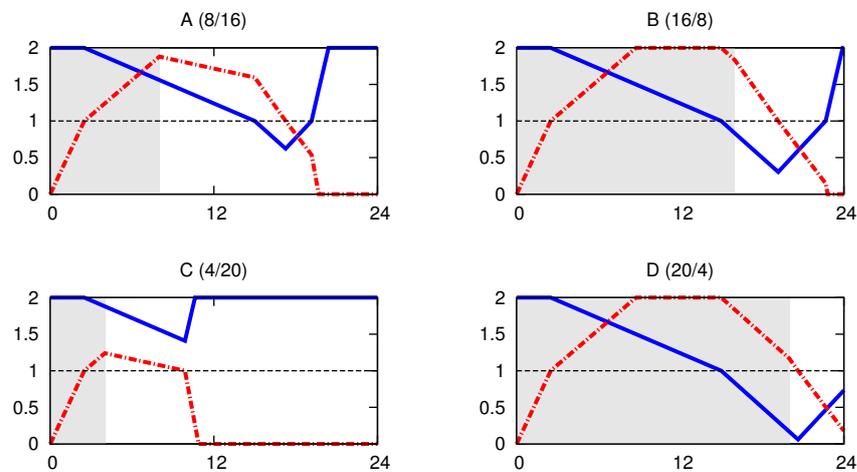
**Figure 14**: Simulations of the model submitted to alternation of dark and light. (A) 8 hours dark followed by 16 hours light. (B) 16 hours dark followed by 8 hours light (C) 4 hours dark followed by 20 hours light (D) 20 hours dark followed by 4 hours light.

It is worth mentioning that our simulations with a remarkably simple model (constant dark condition, summer and winter conditions) are all within the entrainment zone of the onion. At the opposite, the conditions in which the photoperiod is extreme (4 and 20) are not within the onion.

### 5.3  Jet lags

Figure 15 shows the behaviour of the model during a 8 hours jet lag. Figures 15A and 15C represent the behaviour of the model when the circadian clock is affected by a jet lag toward West. Figures 15B and 15D represent its behaviours for a jet lag toward East. Cycles before 48 hours and after 104 (for A and C) and 88 (for B and D) are not represented because they exhibit the same behaviour as in the case of the 12/12 dark/light oscillations of Figure 13C.

- For light phase elongation (A), the first 24 hours are the same as under the equinox condition. In the next 8 hours, the system stays in the waiting state until the next dark phase.

- For the dark phase elongation (C), the first cycle is the same as under the constant dark condition (Figure 13B) where the period is 26 hours, but

the light phase is sufficient to allow the system to reach the waiting state and thus to ensure the synchronisation of the clock to a 24-hours period.
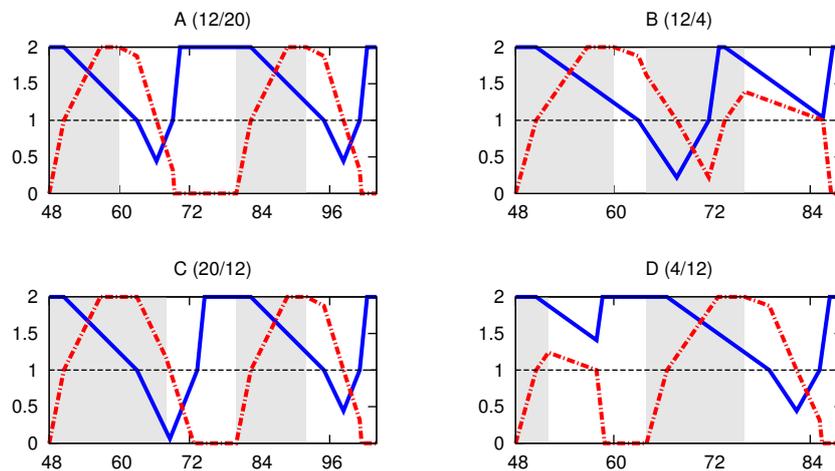


**Figure 15**: Simulations with 8-hours jet lags. (A) Jet lag toward West during the day: The light duration is increased from 12 to 20 hours. (B) Jet lag toward East during the day: The light duration is decreased from 12 to 4 hours. (C) Jet lag toward West during the night: The dark duration is increased from 12 to 20 hours. (D) Jet lag toward East during the night: The dark duration is decreased from 12 to 4 hours. Each simulation is preceded and followed by two 12h dark and 12h light cycles. The windows contain the cycle of the jet lag and the next one.

Nevertheless, the jet lag toward East has more adverse effects.

- When the light phase is shortened (Figure 15B), it can be not long enough to synchronize the clock in the first cycle. Moreover, the shift of the system causes a cycle stop in the next day (the blue curve does not go under 1).

- When the dark phase is shortened (Figure 15D), one observes the same kind of stop in the first cycle but it resumes the next day.

Despite the simplicity of our model, it reproduces notably a lot of real behaviours of the mammalian circadian clock.

## 6 Conclusion

The hybrid formalism presented in this chapter is inspired by the simplicity of the Thomas' modelling framework. The celerities of this modelling framework give the direction and the evolution speed of the system within each discrete state. This formalism mixes the qualitative trajectories as in the Thomas' formalism, with a notion of delays of transitions from a state to another. Such data (delays between 2 distant states) is easily measurable during an experiment.

Proof of concept of our hybrid framework is done through a tiny model of the mammalian circadian clock. The celerities have been identified according to biological behaviours. We first identified constraints on celerities owing to a limit cycle of known period. This method is nevertheless dedicated to small dimensions, so our future work will be dedicated to extensions of the computer aided identification methods already available for the discrete case.

We equipped this new hybrid framework with a simulator generator. This generator allows biologist to generate their own models and make their own *in silico* experiments. This program assists the user to create the network by a web interface, a NetLogo file is then generated and can be executed to observe the behaviour of the model. This interface allows one to choose zeitgebers behaviour (in our example, a scenario for alternations of light and dark) and to modify celerities in real time. Moreover, it shows the level of each variables at each step and the evolution of periods.

Despite its simplicity, our circadian clock model provides interesting behaviours: In constant dark, in 12h/12h alternations of light and dark, in different seasons (modification of the light proportion within a 24-hours period) and when jet lags (modification of one cycle) occur. In constant dark and dark/light alternation conditions, the simulations are very close to the observations. This adaptation faculty is due to the presence of a stable hybrid state (waiting state) allowing the synchronization of the system with the cycle imposed by the external light/dark alternation.

In order to go further on more elaborated models, we have to facilitate the identification of celerities by establishing heuristics able to construct constraints on celerities which have to be satisfied to allow the model to behave as expected. The Hoare logic [4] seems to be a promising first approach.

### *References*

[1] R. Alur, C. Belta, F. Ivančić, V. Kumar, M. Mintz, G. J. Pappas, H. Rubin, and J. Schug. Hybrid modeling and simulation of biomolecular networks. In *Hybrid Systems: Computation and Control*, pages 19–32. Springer, 2001.

[2] D. Barik, W. T. Baumann, M. R. Paul, B. Novak, and J. J. Tyson. A model of yeast cell-cycle regulation based on multisite phosphorylation. *Molecular Systems Biology*, 6(1):405, 2010.

[3] S. Becker-Weimann, J. Wolf, H. Herzel, and A. Kramer. Modeling feedback loops of the mammalian circadian oscillator. *Biophysical Journal*, 87(5):3023 – 3034, 2004.

[4] J. Behaegel, J.-P. Comet, G. Bernot, E. Cornillon, and F. Delaunay. A hybrid model of cell cycle in mammals. *Journal of Bioinformatics and Computational Biology*, 14(1), 2016, In Press.

[5] W. J. Belden, J. J. Loros, and J. C. Dunlap. Clock leaves its mark on histones. *Trends in Biochemical Sciences*, 31(11):610–613, 2006.

[6] G. Bernot and J.-P. Comet. On the use of temporal formal logic to model gene regulatory networks. In *Computational Intelligence Methods for Bioinformatics and Biostatistics*, pages 112–138. Springer, 2010.

[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. *Journal of theoretical biology*, 229(3):339–347, 2004.

[8] K. Bozek, A. Relógio, S. M. Kielbasa, M. Heine, C. Dame, A. Kramer, and H. Herzel. Regulation of clock-controlled genes in mammals. *PLoS One*, 4(3):e4882, 2009.

[9] J.-P. Comet, G. Bernot, A. Das, F. Diener, M. C., and A. Cessieux. *Proc. of the Evry Spring school on Modelling complex biological systems in the context of genomics*, chapter Simplified models for the mammalian circadian clock, pages 85–106. EDP Sciences, 2012.

[10] J.-P. Comet, J. Fromentin, G. Bernot, and O. Roux. A formal model for gene regulatory networks with time delays. In *1st Intl. Conf. on Computational Systems-Biology and Bioinformatics*, pages 1–13. Springer, 2010.

[11] C. Crosio, N. Cermakian, C. D. Allis, and P. Sassone-Corsi. Light induces chromatin modification in cells of the mammalian circadian clock. *Nature neuroscience*, 3(12):1241–1247, 2000.

[12] D. B. Forger and C. S. Peskin. A detailed predictive model of the mammalian circadian clock. *PNAS*, 100(25):14806–14811, 2003.

[13] C. Gérard and A. Goldbeter. Temporal self-organization of the cyclin/cdk network driving the mammalian cell cycle. *PNAS*, 106(51):21643–21648, 2009.

[14] N. R. Glass. Computer analysis of predation energetics in the largemouth bass. *Systems analysis and simulation in ecology*, 1:325–363, 1971.

[15] S. Jamshidi, H. Siebert, and A. Bockmayr. Preservation of dynamic properties in qualitative modeling frameworks for gene regulatory networks. *Biosystems*, 112(2):171–179, 2013.

[16] Z. Khalis, J.-P. Comet, A. Richard, and G. Bernot. The SMBioNet method for discovering models of gene regulatory networks. *Genes, Genomes and Genomics*, 3(1):15–22, 2009.

[17] J.-C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *PNAS*, 100(12):7051–7056, 2003.

[18] C. Schmal, J. Myung, H. Herzel, and G. Bordyugov. A theoretical study on seasonality. *Frontiers in Neurology*, 6(94), 2015.

[19] Y. Shigeyoshi, K. Taguchi, S. Yamamoto, S. Takekida, L. Yan, H. Tei, T. Moriya, S. Shibata, J. J. Loros, J. C. Dunlap, and H. Okamura. Light-induced resetting of a mammalian circadian clock is associated with rapid induction of the mPer1 transcript. *Cell*, 91(7):1043 – 1053, 1997.

[20] E. H. Snoussi and R. Thomas. Logical identification of all steady states: the concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55(5):973–991, 1993.

[21] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42(3):563–585, 1973.

[22] R. Thomas. Regulatory networks seen as asynchronous automata: a logical description. *Journal of Theoretical Biology*, 153(1):1–23, 1991.

[23] S. Troncale, J.-P. Comet, and G. Bernot. Verification of biological models with timed hybrid Petri nets. In *Proc. of the 2007 Intl. Symp. on Computational Models for Life Sciences (CMLS07), ISBN 978-0-7354-0466-3*, volume 952, pages 287–296, Australia, 2007, December 17-19.

[24] K. Vanselow and A. Kramer. Role of phosphorylation in the mammalian circadian clock. In *Cold Spring Harbor symposia on quantitative biology*, volume 72, pages 167–176. Cold Spring Harbor Laboratory Press, 2007.

[25] U. Wilensky. Netlogo. `http://ccl.northwestern.edu/netlogo/`, Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL, 1999.

[26] S.-H. Yoo, S. Yamazaki, P. L. Lowrey, K. Shimomura, C. H. Ko, E. D. Buhr, S. M. Siepka, H.-K. Hong, W. J. Oh, O. J. Yoo, et al. Period2::luciferase real-time reporting of circadian dynamics reveals persistent circadian oscillations in mouse peripheral tissues. *PNAS USA*, 101(15):5339–5346, 2004.

# Towards a Computer Aided Toxicology

Benjamin Miraglio[1], Gilles Bernot[1], Jean-Paul Comet[1],
Christine Risso de Faverney[2]

[1] Université Nice Sophia-Antipolis, CNRS UMR 7271, Laboratoire I3S, France

[2] Université Nice Sophia-Antipolis, CNRS FRE 3729 ECOMERS, Parc Valrose, 28 Avenue Valrose, 06108 Nice, France

### abstract

If the classical paradigm of toxicology has been used for centuries, recent toxicological findings, soaring experimental costs and an increasing regulatory pressure have led the toxicology community towards mechanistic toxicology. This new area of research, focused on molecular envents underlying the toxicity of a chemical substance, motivates the emergence of new modelling approaches for toxicology. In this chapter, we introduce a qualitative rule-based formalism inspired from BioChAM with semantics adapted to the specificities of toxicology. Using a simple example of thyroid hormone system, we then show that this formalism is able to describe the possible toxic disruptions of a biological system. We finally introduce ToxBioNet, a software platform dedicated to toxicology currently under development and we present its already implemented simulator.

## 1   Introduction

The study of adverse effects caused by an exogenous  chemical substance (known as a xenobiotic) in biological systems is called toxicology. The classical toxicology is based on the principle established by Paracelsus in the $XVI^{th}$ century: "All things are poison and nothing is without poison; only the dose makes a thing not a poison." [16] This means that any chemical substance can cause harmful effects to an organism if the system is exposed during a long enough time to a high enough dose of chemical. In modern toxicology, this concept still holds as the basis of the dose-response relationship. In addition, there is almost always a dose below which no response can be measured and conversely, once a maximum response is reached, any increase in the dose will not result in any increased effect. This relationship enables toxicologists to establish a causality between the exposure to a chemical and its induced observed effects. It also allows to determine the threshold of toxicity, namely the lowest exposure (in dose and/or time) where an induced effect occurs.

Many experiments carried out recently have questioned the legitimacy of this paradigm. Indeed, toxicity assessment is quite complex since many factors

can affect the results of toxicity tests. Some of these factors include variables like temperature, food, light, stressful environmental conditions and exposure to other chemical compounds. Other factors related to the test subject itself, including age, sex, health, hormonal status or window of exposure may also greatly influence the vulnerability of an organism to a xenobiotic.

Although lethality is often used to measure toxicity, an increasing trend in toxicology is to focus on the sequence of molecular events occuring during the toxic response and leading to an observable effect. This approach, called mechanistic toxicology, aims to explain the whole causal chain of key events occurring in an organism, from the administration of the compound to its observed adverse effects. In this context, the notion of key event encompasses events occurring at the molecular, cellular and even at the organ scale.

Two almost identical notions concurrently appeared among toxicologists, trying to formalise the chains of key events: the Adverse Outcome Pathways (AOP) [2] and the Modes of Action (MoA) [13]. While only minor parts of their definitions differ, the main distinction between these two notions lies in the context of their use. Indeed, the notion of AOP tends to be used preferentially in ecotoxicology while the MoA notion is mainly used in human toxicology. In this chapter, we only refer to a chain of key events as a pathway of toxicity for the sake of simplicity.

As mechanistic toxicology allows a better understanding of molecular mechanisms leading to adverse effects, it can cope with many difficulties mentioned earlier, such as the extrapolation of toxicity findings obtained from laboratory animals to humans or the consideration of additional factors in toxicity assessments. Moreover, as distinct pathways of toxicity can share the same key events, data obtained when studying one chemical could be reused when assessing other chemicals. By taking all these facts into account, it is very likely that mechanistic bottom-up approaches will complete classical top-down approaches in the near future.

Concurrently, as the potential toxicity of chemical exposure became an area of great concern to both the public [6, 8] and the regulatory authorities [3], the production of chemical compounds is increasingly regulated in the U.S. and in Europe. Manufacturers must now conduct more extensive studies to demonstrate the innocuity of their products, considerably increasing the cost of development of such products.

This context favours the emergence of different modelling approaches, and so far, most of these approaches are quantitative and enable either to infer the toxic threshold of a chemical substance or to confirm its specific pathway of toxicity. To reach these objectives, quantitative approaches need a lot of *in*

*vitro* or *in vivo* toxicological data gathered during the early stages of the development process of the chemical substance. This necessity can be restrictive given the current cost of acquiring new biological data. There is therefore an incentive to develop new approaches that do not focus on toxic thresholds. Instead, they aim to describe pathways of toxicity at the *qualitative* level, namely by discretising continuous concentrations into intervals of interest. These approaches try to enumerate all the possible pathways of toxicity included in a biological system and then check the biological plausibility of these pathways. Their final goal is to highlight the most probable pathways involved in a given toxicity.

In this chapter, we present a new qualitative formalism allowing to describe a biological system with its possible toxicological disruptions. This formalism was originally inspired from the boolean semantics of BioChAM [5], an environment able to model biological systems as networks of chemical reactions. However, these semantics are somehow too rough for toxicology due to particular features present in the toxicological models. Our new formalism therefore extends the boolean semantics of BioChAM to take into account these specificities, such as the notion of abnormal concentrations or the presence of modulating interactions impossible to manage similarly to classical chemical reactions. The purpose of this new formalism is to help toxicologists in their search for new pathways of toxicity.

Throughout this chapter, the presentation of the formalism will be illustrated by the thyroid hormone system. Indeed, this system is one of the least sex hormone dependent system and its mechanisms are well described in the literature. The next section is thus dedicated to the description of the thyroid hormone system and the various mechanisms ensuring its homeostasis. In Section 3, we explain how to use the new formalism to construct a toxicological system and the associated semantics. This formalism is then applied to a simplified thyroid hormone system in Section 4 and finally, we describe the aim of the ToxBioNet software platform and its first component, a simulator dedicated to our formalism, in Section 5.

## *2  Thyroid Hormone Homeostasis*

The underlying biological network ensuring the homeostasis of the thyroid system is complex and results in a finely regulated system where thyroid hormone levels only vary subtly during the day [19]. The homeostasis of thyroid system is necessary since any perturbation of this system can have major effects on the health of individuals, especially when it occurs in the earliest stages of development of an organism [1].

***The hypothalamo-pituitary-thyroid axis (HPT axis)*** is part of the neuroendocrine system involved in the regulation of metabolism and in the thyroid homeostasis in particular. As suggested by its name, this axis is composed of three compartments: the hypothalamus, the pituitary and the thyroid gland. The hypothalamus is a brain structure that controls endocrine glands. Part of this region secretes a neuropeptide, Thyrotropin-Releasing-Hormone (TRH). TRH is transported in axonal fluid to stimulate thyrotrophic cells in the anterior pituitary gland, stimulation that triggers the synthesis and secretion of Thyroid Stimulating Hormone (TSH). TSH is released into blood circulation and stimulates the follicular cells of the thyroid gland, leading to the synthesis of thyroid hormones (TH) and their secretion into the blood circulation [15].

***Thyroid hormones (TH)*** are derived from the tyrosine amino acid and can be iodinated at different levels. For example, tri-iodo-thyronine (T3) and tetra-iodo-thyronine, also known as thyroxine (T4), are respectively iodinated three and four times. Moreover, the position of iodine residues in the chemical structure is important for the function of the hormone. Indeed, the reverse tri-iodo-thyronine (rT3) is as iodinated as T3 but does not have the same effects since its iodine residues are not located in the same places.

Historically, T3 is considered as the sole active form of thyroid hormone, T4 only being a pro-hormone that can be activated into T3 by deiodination [9]. Most of the T4 is converted into T3 in the liver. In this classic view, the action of TH on target genes is mediated by Thyroid hormone Receptors (TR). These receptors are constitutively located in the cell *nucleus* of any cell targeted by the thyroid hormone. TR can bind to T3 and more marginally to T4 [23]. While TR-T4 complexes are ineffective [23], TR-T3 complexes present the ability to bind to precise regions of DNA called thyroid hormone response elements. Once binded to these elements, TR-T3 complexes can then influence the transcription of target genes, either in a positive or a negative manner depending on the gene [23].

Recent studies have shown that T4, rT3 and other products of TH deiodination also have a biological activity that does not involve TR [14, 20]. These actions are currently under further investigations by the endocrinology community and will not be developed in this chapter.

***Several negative feedbacks*** are present in the HPT axis in order to ensure a proper regulation of the thyroid system. Actually, the production of TRH and TSH are repressed by the negative feedback effects of T3 over respectively the hypothalamus and the anterior pituitary [10, 7]. T3 also stimulates the production of Pyroglutamyl Peptidase 2 (PP2), an enzyme known to destroy

the TRH before it can reach the pituitary gland [18]. Those regulatory effects are mediated by the binding of T3 to TR located in target cells.

***Degradation of TH*** is ensured by two types of enzymes. Specific enzymes such as the deiodinases (Dio), present in different tissues such as the liver, the kidney or the brain, inactivate TH by removing one or more iodine residues [11]. In contrast, the non-specific hepatic enzymes such as the glucuronosyltrans-ferases (UGT) conjugate TH with a residue leading to the biliary excretion of TH [17]. It should be noted that these non-specific hepatic enzymes are also generally activated when the organism is exposed to some foreign chemicals such as drugs or poisons considered as *indirect Endocrine Disruptors (iED)*. In this context, an over-activation of UGT can cause a major decrease in blood TH and disrupt indirectly the thyroid hormone homeostasis [17].

Other foreign chemicals, considered as *direct Endocrine Disruptors (dED)* can also disrupt the thyroid hormone homeostasis by acting directly on the production of TH in the thyroid gland [12]. Thus, a direct or indirect disruption leading to the decrease of TH blood levels will result in the compensatory increase of both TRH and/or TSH levels. This increase, when it is lasting, may raise the risk of developping some thyroid cancers [4, 12].

***We propose at first a simple representation of the thyroid hormone system*** that can summarise this knowledge about the physiology of the thyroid hormone system. This representation comprises TRH, TSH, TH, TR, PP2, Dio, UGT and we can also add a direct endocrine disruptor (abbreviated as ED for the sake of simplicity) that will interfere with the production of TH (see Figure 1).

For pedagogical purposes, we will over-simplify this representation. Here, since TRH only stimulates TSH production, we can abstract it and assume that PP2 and TR-TH negative interactions directly concern TSH. Since PP2 and TR-TH negative influences both originate from TH and concern both TSH, we can also abstract PP2. Moreover, since the ED only disrupts directly TH levels, we can put aside the TH degradation processes and pull out Dio and UGT from the representation. In the end, the resulting simplified representation only includes TH, TSH, TR and the ED (see Figure 2).

## 3   A New Discrete Framework for Toxicology

A biological system can be described as a set of biological entities interacting with each other at different concentrations. For each entity, there exists a concentration regarded as normal in standard conditions in a given organism. For instance, in an adult human, the normal blood concentration of glucose is 1 g/L.
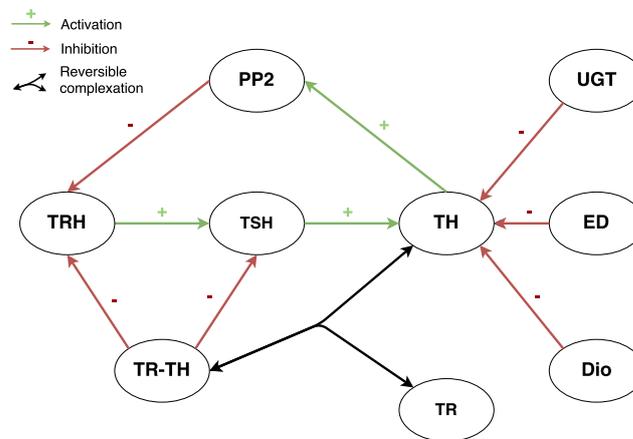
**Figure 1**: Simple representation of the thyroid hormone system.
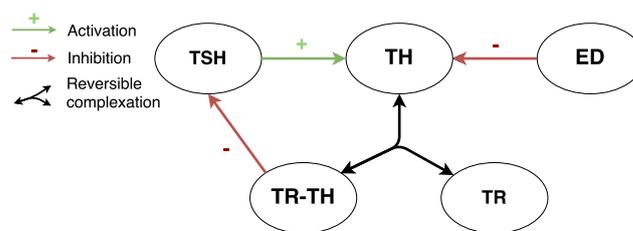


**Figure 2**: Over-simplified representation of the thyroid hormone system.

In order to represent the evolution of the concentration of each entity and to detect abnormal concentrations, we introduce four qualitative abstract levels, which are enumerated here in increasing order:

- $\varepsilon$ reflects a negligible concentration of a given entity, that is to say a concentration too low to trigger any mechanism in the biological system.

- $\iota$ conveys an abnormally low concentration, *i.e.* a relative lack of this entity that can affect some mechanisms in the biological system.

- $\Delta$ indicates a normal concentration.

- $\theta$ shows an abnormally high concentration, namely an excess of this entity.

In a given biological system, not all entities have abnormally low or high concentrations depending on the studied issue. Therefore in this formalism, only the levels $\varepsilon$ and $\Delta$ are mandatory for each entity, $\iota$ and $\theta$ are optional. All these facts are gathered in the signature of a biological system which defines the set of biological entities present in the system and the levels admissible for each entity.

**Definition 1 [Signature]** *A* signature *is a finite set $E$ whose elements are named entities. Moreover, $E$ is given an application $\tau : E \to \mathcal{P}(\mathbb{L})$ where $\mathbb{L} : \{\varepsilon, \iota, \Delta, \theta\}$, and such that for each entity $e \in E$, $\{\varepsilon, \Delta\} \subset \tau(e)$. $\tau(e)$ is the set of admissible levels of the entities $e$. Moreover, by convention, $\mathbb{L}$ is equipped with the strict total order relation $\varepsilon < \iota < \Delta < \theta$.*

For instance, the signature of the simplified thyroid hormone system corresponds to the set of five entities {TSH, TH, TR, TR-TH, ED} and each entity has its own set of admissible levels. For example the set of admissible levels of TH, $\tau(\text{TH})$, is equal to $\{\varepsilon, \iota, \Delta, \theta\}$ as TH can be in excess or abnormally low in some cases.

Once the system signature is defined, it is then possible to define the state of the system as the qualitative levels of all entities of the system. For example the simplified thyroid hormone system can be at a state $\eta_0$ where TSH is at the level $\theta$, noted $\eta_0(\text{TSH}) = \theta$ and where $\eta_0(\text{TH}) = \iota$, $\eta_0(\text{TR}) = \Delta$, $\eta_0(\text{TR-TH}) = \iota$ and $\eta_0(\text{ED}) = \varepsilon$. This state can then be written:

$$\eta_0 = (\theta, \iota, \Delta, \iota, \varepsilon) \quad (1)$$

where the order of variable is TSH, TH, TR, TR-TH, ED.

**Definition 2 [State]** *A signature $E$ being given, a* state *$\eta$ is a function $E \to \mathbb{L}$ such that for all $e \in E$, $\eta(e) \in \tau(e)$.*

In order to represent the evolution of the system, we introduce two functions: the incrementation, noted $incr$, and the decrementation, noted $decr$. These functions apply to one entity at a time and return the level of this entity just above (resp. below) its current level. Because all entities have not the same set of admissible levels, there is one function defined for each entity. For instance, if $\tau(\text{TSH}) = \{\varepsilon, \Delta, \theta\}$ and $\eta_0(\text{TSH}) = \Delta$, then $incr_{\text{TSH}}(\eta_0(\text{TSH})) = \theta$ and $decr_{\text{TSH}}(\eta_0(\text{TSH})) = \varepsilon$. It should be noted that the incrementation (resp. decrementation) function is not defined on the maximal (resp. minimal) level of the admissible levels. Therefore, in our previous example, $incr_{\text{TSH}}(\eta(\text{TSH}))$ is not defined if $\eta(\text{TSH}) = \theta$.

Alongside these functions, some properties on the entity levels can be described by formulas.

**Definition 3 [Formula]** *A signature $E$ being given, the set of formulas on $E$ is inductively defined by:*

- *for all symbols $a$ and $b$ belonging to $E$ or $\mathbb{L}$, the atoms $a = b$, $a > b$, $a \geqslant b$, $a \leqslant b$ and $a < b$ are atomic formulas.*

- *if $\varphi$ and $\psi$ are well-formed formulas on $E$, then $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \Rightarrow \psi$ are also well-formed formulas on $E$.*

**Definition 4 [Satisfaction relation]** *A state $\eta$ and a formula $\varphi$ on a signature $E$ being given, the satisfaction relation $\eta \vDash \varphi$ is inductively defined by:*

- *if $\varphi$ is an atom of the form $a = b$, then $\eta \vDash \varphi$ if and only if $\overline{\eta}(a) = \overline{\eta}(b)$ where $\overline{\eta}$ is the extension of $\eta$ to $E \cup \mathbb{L}$ by the identity on $\mathbb{L}$. We proceed similarly for the other comparison predicates.*

- *if $\varphi$ is of the form $\varphi_1 \wedge \varphi_2$ then $\eta \vDash (\varphi_1 \wedge \varphi_2)$ if and only if $\eta \vDash \varphi_1$ and $\eta \vDash \varphi_2$. We proceed similarly for the other connectives.*

For instance, the formula $\varphi$ stating the presence of TR at a normal level can be written as: TR $= \Delta$ and the formula $\psi$ stating that the level of TH is strictly superior to the one of TSH can be written as: TH $>$ TSH. The state $\eta_0$, previously described in eq. 1, satisfies $\varphi$ but not $\psi$.

To determine the evolution of the biological system, a set of rules is given. This set is interpreted as biological transformations. In short, a rule can be resumed by the following representation:

$$A_1 + \cdots + A_m \quad \Rightarrow \quad A_{m+1} + \cdots + A_n \ \ up(\varphi) \ \ down(\psi)$$

Each rule includes two sets of entities, the first one, for all $i$ in $[1, m]$, constitutes the set of "reactants" while the other one, for all $i$ in $[m+1, n]$, represents the set of "products." A rule also includes two modulating conditions $up(\varphi)$ and $down(\psi)$ ($\varphi$ and $\psi$ being formulas) representing respectively a positive and a negative possible modulation of the rule. The $up(\varphi)$ (resp. $down(\psi)$) modulation takes only effect if $\varphi$ (resp. $\psi$) is satisfied and its effects are further detailed later on. Of course, if no modulation is known for a given rule, it is not displayed in the rule representation.

**Definition 5 [Biological action network]** *A biological action network on a signature $E$ is a set of rules where each rule is an expression of the form:*
$$A_1 + \cdots + A_m \quad \Rightarrow \quad A_{m+1} + \cdots + A_n \ \ up(\varphi) \ \ down(\psi)$$
*where:*

- *$\forall i \in \{1 \ldots n\}, A_i \in E$.*

- *$\varphi$ and $\psi$ are formulas on $E$.*

Notice that a rule can be devoid of any reactant or product. In the previous definition, the index $m$ can be equal to zero (the rule does not need any reactant) or $m$ can be equal to $n$ (the rule has no product). A rule without reactant can be considered as the constitutive production of an entity in a given model and a rule without product can be interpreted as the degradation of an entity. In either cases, the empty solution is depicted using the _ symbol.

It is worth mentioning that despite the strong resemblance between a rule and a chemical reaction, a rule must *not* be interpreted as quanta of reactants converted into quanta of products but as a possible evolution of levels of entities present in the rule.

As a basic example of rule, the complexation of TH with TR can be represented by the following rule:

$$\text{TH} + \text{TR} \quad \Rightarrow \quad \text{TR-TH}$$
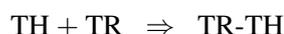
Since neither positive nor negative modulating conditions are considered here, only reactants and products are displayed.

In order to be applicable at a given state, a rule must meet basic criteria inspired from biology. First, since the level $\varepsilon$ is interpreted as a negligible concentration, a rule is applicable only if all its reactants are present at least at the level $\iota$. In addition, a rule cannot be applied if the negative regulating condition $down()$ applies, namely if the corresponding formula is satisfied.

**Definition 6 [Applicable rule]** *Let $\eta$ be a state and let us consider a rule $r$ of the form $A_1 + \cdots + A_m \Rightarrow A_{m+1} + \cdots + A_n \; up(\varphi) \; down(\psi)$. The rule $r$ is* applicable *at the state $\eta$ if and only if:*

- $\forall i = 1 \ldots m, \; \eta(A_i) \neq \varepsilon$.

- $\eta \nvDash \psi$.

For instance, let us consider the deiodination of T4 into T3 by the type 2 deiodinase (dio2). If we assume that $\tau(\text{dio2}) = \{\varepsilon, \iota, \Delta, \theta\}$, the deiodination can be written as:

$$\text{T4} \quad \Rightarrow \quad \text{T3} \; down(\text{dio2} < \Delta) \quad (a)$$
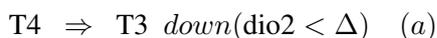
This rule is applicable if and only if the level of T4 is strictly greater than $\varepsilon$ and the level of type 2 deiodinase is at least $\Delta$, namely if there is T4 in the system and a normal concentration of type 2 deiodinase. Note that the catalysis, namely the necessary presence of an enzyme to the proper conduct of a reaction, can be expressed using the $down()$ condition as in the previous example.

When a rule is applied, part of its entities can vary to a potential next level. Since a reactant is consumed during the application of a rule, it is possible for

its level to cross a downward threshold and become lower than its initial value. Therefore, the next level of a reactant is the one returned by the decrementation function applied to that reactant, or the current level if the threshold is not crossed. The next potential level of a product is determined by the levels of reactants participating in the rule. The idea is simple: a product can increase only if each reactant is at a level sufficient to allow the considered product to increase. This is represented here by the condition that the level of every reactant must be strictly greater that the level of the product. Therefore, the next potential level of a product is returned by the incrementation function applied to it only if the level of every reactant in the rule is strictly greater than the initial level of the product.

The notable exception to this qualitative evaluation is the over-activated rules, namely, rules where the $up()$ condition applies. If a rule is over-activated, then the next potential level of a product is always returned by the incrementation function applied to it, independently of the reactant levels.

**Definition 7 [Potential next level]** *Let $\eta$ be a state and $r$ be a rule of the form $A_1 + \cdots + A_m \Rightarrow A_{m+1} + \cdots + A_n \ \ up(\varphi) \ \ down(\psi)$, applicable in $\eta$,*

- *for each reactant $R \in \{A_1 \ldots A_m\}$, the potential next level of $R$ by $r$ is $decr_R(\eta(R))$.*

- *if $\eta \vDash \varphi$, then for each product $P \in \{A_{m+1} \ldots A_n\}$, the potential next level of $P$ by $r$ is $incr_P(\eta(P))$.*

- *if $\eta \nvDash \varphi$, then for each product $P \in \{A_{m+1} \ldots A_n\}$, the potential next level of $P$ by $r$ is $incr_P(\eta(P))$ only if*
$$\eta(P) < \min_{R \in \{A_1 \ldots A_m\}} (\eta(R)).$$
*with $\min\limits_{R \in \{\}} (\eta(R)) = \Delta$.*

Notice that the constitutive production of an entity, represented by a rule devoid of any reactant, does not inherently lead to an abnormally high level of the entity. Therefore, when the set $R$ of reactant is empty, its minimum is considered to be $\Delta$.

The restriction on the possible evolution of product levels (third item of definition 7) relies on the assumption that the levels of entities do not spontaneously evolve towards anormal conditions but, indeed, need an initiating factor such as a pre-existing disorder in the reactant levels or the over-activation of the rule to reach abnormal levels. If we keep the T4 deiodination as an example, we can also specify that an excess of type$\tilde{2}$ deiodinase can cause trouble in T3 levels by adding a $up()$ condition to the rule $(a)$:

$$\text{T4} \Rightarrow \text{T3} \ down(\text{dio2} < \Delta) \ up(\text{dio2} > \Delta)$$

Here, assuming that the rule is applicable at the state $\eta_0$ and that $\eta_0(\text{T3}) = \Delta$, the potential next level of T3 by this rule can be $\theta$ only if $\eta_0(\text{T4}) = \theta$ or if $\eta_0(\text{dio2}) > \delta$ (so, $\eta_0(\text{dio2}) = \theta$).

Among all the applicable rules at a given state, only one is applied at a time. When a rule is applied, one and only one of its entities evolves to its potential next level. This means that the level of an entity has to change in order to consider that the rule was applied. Importantly, this also means that neither reactant nor product levels are updated simultaneously. Similar ideas have been firstly developed for discrete gene models by Thomas and Snoussi [21, 22]. This behaviour reflects the possibility for the level of a reactant to cross a threshold without all the other reactant levels having to also cross a threshold.

In brief, starting from a given state, it is possible to determine which rules of the system are applicable at that state. Among these rules, the application of one rule changes the level of one entity, modifying the system state. It is then possible to establish a transition graph, mapping all the possible transitions between the states of a system.

**Definition 8 [Transition graph]** *A biological action network $N$ being given, the associated* transition graph *is the graph $G = (V, T)$ whose set $V$ of vertices is the set of states on the signature $E$ of $N$, and such that there exists an edge from a state $\eta$ to a state $\eta'$, called transition and noted $\eta \to \eta'$, if and only if:*

- *there exists a rule $r$ of the form $A_1 + \cdots + A_m \Rightarrow A_{m+1} + \cdots + A_n$ $up(\varphi) \ down(\psi)$ applicable at $\eta$.*

- *there exists a unique index $i \in [1 \ldots n]$ such thaot the potential next level of $A_i$ by $r$ is $\eta'(A_i)$ and $\forall e \in E \smallsetminus \{A_i\}$, $\eta'(e) = \eta(e)$. In other words,o the only changed level is the level of $A_i$ becoming $\eta'(A_i)$.*

In fact, it is possible to loop on an state even if there is an outgoing transition. This means that self loops are present on every state but they are not included here to avoid an overburden of the transition graph.

Once the transition graph of the biological system is established, it can be used as a basis for testing properties about the system dynamics.

### 4  Application of the Formalism to the Thyroid Hormone System

According to Figure 2, the signature of the system is the set of entities {TSH, TH, TR, TR-TH, ED}. The set of admissible levels of each entity is determined according to the rules where this entity intervenes, so we first detail the rules of the system representing the different interactions presented in Figure 2:

1. $\_ \Rightarrow$ TR

2. TR + TH $\Rightarrow$ TR-TH

3. TR-TH $\Rightarrow$ TR + TH

4. $\_ \Rightarrow$ TSH $down(\text{TR-TH} \geqslant \Delta)$ $up(\text{TR-TH} = \varepsilon)$

5. TSH $\Rightarrow$ TH $down(\text{ED} > \varepsilon)$ $up(\text{TSH} > \Delta)$

Since TR receptors are formed constitutively in tissues sensitive to TH, the first rule abstracts the production of TR. Rules 2 and 3 represent respectively the complexation of TR and TH into TR-TH and their decomplexation. As a reminder, TR-TH is the entity that will determine the negative feedback strength applied on TSH production.

Rule 4 abstracts the molecular machinery allowing the production of TSH. This machinery is inhibited by the presence of TR-TH (in accordance with the negative feedbacks paragraph in Section 2), via the $down()$ condition. Conversely, when the TR-TH concentration is insignificant, the production of TSH is over-activated in a compensating effort by the organism, as formalised by the $up()$ condition. So, TSH is considered to be produced normally only when the concentration of TR-TH is at an abnormally low level but still significant in the organism, namely $\iota$.

Finally, Rule 5 represents the TH production preconditioned by the level of TSH. TSH is considered as a reactant to take into account the inherent degradation of TSH during the TH production. The $down()$ condition introduces the endocrine disruptor action that blocks TH production and the $up()$ condition allows for the possibility of TH reaching an abnormally high concentration when over-stimulated by TSH. These rules induce a precise definition of the different sets of admissible levels:

- Since TH is the main concern of the model, its level should be as accurate as possible, therefore $\tau(\text{TH}) = \{\varepsilon, \iota, \Delta, \theta\}$.

- In order to allow an over-activation of the TH production (rule 5), the TSH level must be able to reach $\theta$. On the contrary, $\iota$ is not required here, thus $\tau(\text{TSH}) = \{\varepsilon, \Delta, \theta\}$.

- The same applies to the TR-TH set of admissible level: $\iota$ is necessary to allow a normal production of TSH (rule 4) but not $\theta$. Therefore, $\tau(\text{TR-TH}) = \{\varepsilon, \iota, \Delta\}$.

- In this model, we are only interested in the presence or absence of TR and ED. We can then assimilate their levels to Boolean values: $\tau(\text{TR}) = \{\varepsilon, \Delta\}$ and $\tau(\text{ED}) = \{\varepsilon, \Delta\}$.

The complete graph has five dimensions and includes 144 different states ($144 = 4 \times 3 \times 3 \times 2 \times 2$). Here we focus only on the specific region of this graph where the TR level is $\Delta$ (*i.e.* we consider it constitutively expressed) and the ED level is $\varepsilon$ (because we first want to see the behaviour of the system when not disrupted). These restrictions limit the graph to 36 states (represented in three dimensions) and make it representable as in Figure 3.

It is easily observable that the only way to reach the plane where $\eta(\text{TH}) = \theta$ is to go through the dashed green arrows present on the plane $\eta(\text{TSH}) = \theta$. These arrows represent the transitions allowed by an over-activation of the production of TH (rule 5) when TSH is in excess. Furthermore, the plane $\eta(\text{TSH}) = \theta$ is only reachable by the dashed red arrows on the the plane $\eta(\text{TR-TH}) = \varepsilon$ corresponding to the over-activation of the production of TSH (rule 4). Finally, we also see that in order to reach the plane $\eta(\text{TR-TH}) = \Delta$, the predecessor state must have a TH level of at least $\Delta$, illustrating Definition 7.

If we introduce an endocrine disruptor to the system and thus put the level of ED to $\Delta$, we can see that all the green arrows would disappear because of the $down()$ condition on the production of TH (rule 5). At this point, all the transitions converge towards the state $\theta\varepsilon\varepsilon$ at the bottom left in the background. This state corresponds to an excess of TSH combined to a lack of TH and TR-TH, namely the same condition as observed biologically when an endocrine disruptor is introduced into an organism.

This example, although simple, shows that the abstraction made by this formalism is adequate to reproduce known biological behaviours.

## 5  ToxBioNet, a Software Platform Dedicated to Toxicology

The aim of ToxBioNet is to help toxicologists in their search for potential toxic pathways explaining the mechanism of a toxicity. One of the main goals is to be able to extract all the possible pathways between an initiating event and an adverse outcome, compatible with a particular biological action network. The resulting pathways can be further filtered if some key events are experimentally known to be involved or not involved in the studied toxicity. Similarly, some successions of key events are known to be highly unlikely in biology (for example an event A known to be never followed by the event B). Pathways including such successions can thus be also filtered out. Finally, we would like to develop a heuristic suggesting the most informative and relevant experiences when trying to determine which remaining pathways are actually involved in the studied toxicity. To achieve this purpose, our formalism has to be combined with a second language able to express properties such as successions of key events.
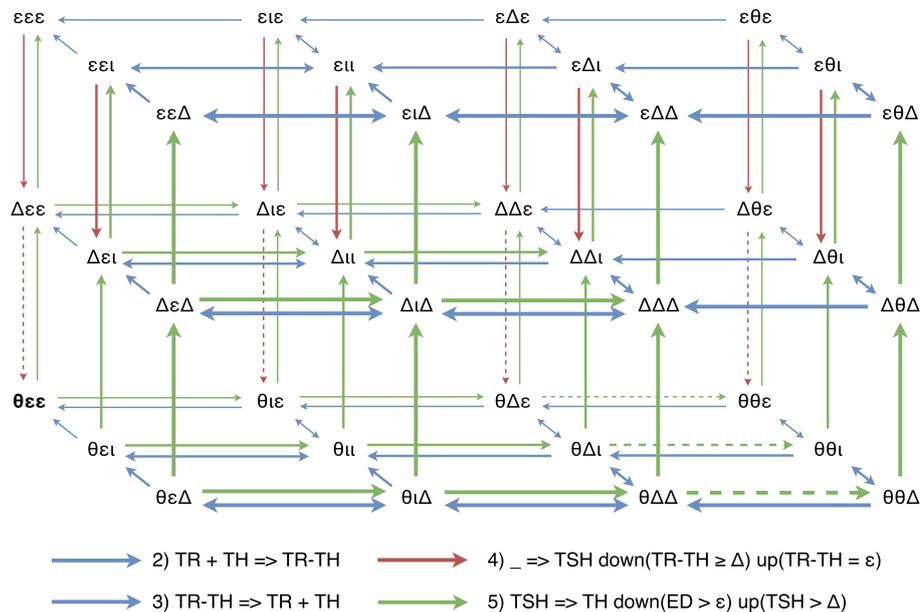
**Figure 3**: Partial transition graph of the over-simplified thyroid hormone model. The states are represented by a 3-letters string where the first one (resp. second and third) is the level of TSH (resp. TH and TR-TH). The levels of TR and ED are set to $\Delta$ and $\varepsilon$ respectively. A dashed arrow indicates the application of an over-activated rule, where the $up()$ condition applies.

Formal methods will be helpful to assist toxicologists to construct a mechanistic model. These can reveal previously unsuspected relations between pathways or key events. When trying to enumerate the pathways leading from an event $A$ to an event $B$, the filtering step is facilitated by the search for inconsistencies between existing knowledge and hypotheses. As a trivial example, if an event of a pathway $P$ from $A$ to $B$ is involved in another pathway which is certain to lead to $B'$, and if we know that $B$ never happens alongside $B'$, the pathway $P$ can be easily filtered out.

ToxBioNet is currently under development: it is already possible to run simulations on biological action networks. This simulator, written in Java, is able to parse an input file containing all the rules describing a biological system and to create the corresponding toxicological model. As shown in Section 4, the number of states in the transition graph grows exponentially with the number of entities (a system including $n$ entities can have up to $4^n$ states). This makes the generation of the state graph technically difficult if not impossible

for systems including more than fifteen entities (which implies approximately a billion possible states). In order to avoid the pitfall of the construction of a huge state graph, the simulator can exhibit as many traces as wanted in the state graph without constructing it. It was tested using a complex thyroid hormone system model: this model includes more than fifty rules based on a hundred of scientific references.

## 6   Conclusion

In this chapter is presented a new formal framework able to handle several specificities of toxicology not taken into account so far, such as the possible presence of a compound in abnormal concentrations or the possibility, for a reaction, to be modulated. This modelling framework is applied to the simple model of the thyroid hormone system and its expressive power allows us to describe the biological system with enough precision to reproduce existing behaviours such as the disruption of TH levels resulting in abnormally high TSH levels.

In the future, the current formalism will be combined with a formal language able to express properties on successions of key events: a simple idea would be to adapt classical temporal logics to our toxicological framework. Formal methods such as model checking, will be useful to find new potential pathways of toxicity satisfying some given properties. Besides, checking the successions of key events could also highlight gaps in the current toxicological knowledge. The platform ToxBioNet will therefore be useful as an experiment-aid tool to select the most informative experiments to conduct.

## References

[1] S. Andersen, K. M. Pedersen, N. H. Bruun, and P. Laurberg. Narrow individual variations in serum t4 and t3 in normal subjects: a clue to the understanding of subclinical thyroid disease. *The Journal of Clinical Endocrinology & Metabolism*, 87(3):1068–1072, 2002.

[2] G. T. Ankley, R. S. Bennett, R. J. Erickson, D. J. Hoff, M. W. Hornung, R. D. Johnson, D. R. Mount, J. W. Nichols, C. L. Russom, P. K. Schmieder, et al. Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry*, 29(3):730–741, 2010.

[3] T. Backhaus, H. Blanck, and M. Faust. *Hazard and risk assessment of chemical mixtures under REACH: state of the art, gaps and options for improvement*. Swedish chemicals agency, 2010.

[4] R. A. Barter and C. D. Klaassen. Reduction of thyroid hormone levels and alteration of thyroid function by four representative udp-glucuronosyltransferase inducers in rats. *Toxicology and applied pharmacology*, 128(1):9–17, 1994.

[5] L. Calzone, F. Fages, and S. Soliman. Biocham: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics*, 22(14):1805–1807, 2006.

[6] T. Colborn, D. Dumanoski, and J. P. Myers. *Our stolen future: are we threatening our fertility, intelligence, and survival?: a scientific detective story*. Dutton, 1996.

[7] R. H. Costa-e Sousa and A. N. Hollenberg. Minireview: the neural regulation of the hypothalamic-pituitary-thyroid axis. *Endocrinology*, 153(9):4128–4135, 2012.

[8] J. Kepner. Synergy: The big unknowns of pesticide exposure. *Pesticides and You*, 23(4):17–20, 2004.

[9] J. Köhrle. Local activation and inactivation of thyroid hormones: the deiodinase family. *Molecular and cellular endocrinology*, 151(1):103–119, 1999.

[10] F. Lezoualc'h, A. H. Hassan, P. Giraud, J.-P. Loeffler, S. L. Lee, and B. A. Demeneix. Assignment of the beta-thyroid hormone receptor to 3, 5, 3'-triiodothyronine-dependent inhibition of transcription from the thyrotropin-releasing hormone promoter in chick hypothalamic neurons. *Molecular Endocrinology*, 6(11):1797–1804, 1992.

[11] R. Malik and H. Hodgson. The relationship between the thyroid gland and the liver. *Qjm*, 95(9):559–569, 2002.

[12] P. Männistö, T. Ranta, and J. Leppäluoto. Effects of methylmercaptoimidazole (mmi), propylthiouracil (ptu), potassium perchlorate (kclo4) and potassium iodide (ki) on the serum concentrations of thyrotrophin (tsh) and thyroid hormones in the rat. *Acta endocrinologica*, 91(2):271–281, 1979.

[13] M. E. Meek, A. Boobis, I. Cote, V. Dellarco, G. Fotakis, S. H. A. R. O. N. Munn, J. Seed, and C. Vickers. New developments in the evolution and application of the who/ipcs framework on mode of action/species concordance analysis. *Journal of Applied Toxicology*, 34(1):1–18, 2014.

[14] L. C. Moeller, X. Cao, A. M. Dumitrescu, H. Seo, and S. Refetoff. Thyroid hormone mediated changes in gene expression can be initiated by cytosolic action of the thyroid hormone receptor $\beta$ through the phosphatidylinositol 3-kinase pathway. *Nuclear receptor signaling*, 4:e020, 2006.

[15] OECD. In vitro & ex vivo assays for identification of modulators of thyroid hormone signalling. Technical report, Organisation for Economic Cooperation and Development, 2013.

[16] Paracelsus. *Das Buch Paragranum / Septem Defensiones*. Hofenberg, 2014.

[17] K. Saito, H. Kaneko, K. Sato, A. Yoshitake, and H. Yamada. Hepatic udp-glucuronyltransferase (s) activity toward thyroid hormones in rats: induction and effects on serum thyroid hormone levels following treatment with various enzyme inducers. *Toxicology and applied pharmacology*, 111(1):99–106, 1991.

[18] E. Sánchez, M. A. Vargas, P. S. Singru, I. Pascual, F. Romero, C. Fekete, J.-L. Charli, and R. M. Lechan. Tanycyte pyroglutamyl peptidase ii contributes to regulation of the hypothalamic-pituitary-thyroid axis through glial-axonal associations in the median eminence. *Endocrinology*, 150(5):2283–2291, 2009.

[19] U. Schweizer, J. M. Weitzel, and L. Schomburg. Think globally: act locally: new insights into the local regulation of thyroid hormone availability challenge long accepted dogmas. *Molecular and cellular endocrinology*, 289(1):1–9, 2008.

[20] A. I. Shih, S. Zhang, H. J. Cao, H.-Y. Tang, F. B. Davis, P. J. Davis, and H.-Y. Lin. Disparate effects of thyroid hormone on actions of epidermal growth factor and transforming growth factor-$\alpha$ are mediated by 3*, 5*-cyclic adenosine 5*-monophosphate-dependent protein kinase ii. *Endocrinology*, 145(4):1708–1717, 2004.

[21] E. H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. *Dynamics and stability of Systems*, 4(3-4):565–583, 1989.

[22] R. Thomas. Regulatory networks seen as asynchronous automata: a logical description. *Journal of theoretical biology*, 153(1):1–23, 1991.

[23] P. M. Yen, S. Ando, X. Feng, Y. Liu, P. Maruvada, and X. Xia. Thyroid hormone action at the cellular, genomic and target gene levels. *Molecular and cellular endocrinology*, 246(1):121–127, 2006.

# Verilog-A Compact Space-dependent Model for Biology

Elise Rosati[1], Morgan Madec[1], Jean-Baptiste Kammerer[1], Abir Rezgui[1],
Christophe Lallement[1], Jacques Haiech[2]

[1] Engineering Science, Computer Science and Imaging Laboratory (ICube,
   UMR 7357), Bd Sébastien Brandt, F-67412 Illkirch Cedex, France
[2] Therapeutic Innovation Laboratory (LIT, UMR 7200) Faculté de Pharmacie
   74 route du Rhin, F-67400 Illkirch Cedex, France

## *Abstract*

We recently demonstrated that it is possible to model and to simulate biological functions using hardware description languages (and associated simulators) traditionally used for micro-electronics. The main drawback of these languages is that they do not support partial differential equations. However, for several applications in biology, space-dependent quantities are unavoidable. This paper deals with a new approach to address these problems. Our work is based on previous investigations on electro-thermal simulations in integrated circuits. The tool is composed of four main parts: a mesher that divides space into small cubes (or squares in 2D), a set of interconnected biological models, a SPICE simulator that handles these models and a Python script that interfaces the different tools. Simulation results obtained with our tool are compared with experimental data for a specific case. Results are in good accordance from a qualitative viewpoint.

*Keywords – Compact modeling; Verilog-A; Space-and-Time modeling; mesher; sytems and synthetic biology*

## 1  Introduction to Synthetic Biology

Synthetic Biology is defined by European experts as "the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature. [...] In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way" (Synthetic Biology: Applying Engineering to Biology: Report of a NEST High Level Expert Group). By designing biological networks and new molecules, biologists and engineers want to achieve computing-like behaviors in cells by rewiring and reprogramming them [1]. Synthetic biology encompasses focuses ranging from genetic circuit design [2] and synthesis of engineered proteins [3] to metabolic engineering [4] and minimal cell or protocell [5]. Synthetic constructions allowed advances in fundamental research and are now being

applied to solve concrete problems in various domains (biosensing, therapeutics, biofuels [6], novel biomaterials [1]).

Because of their similarities to electronic circuits, our main focus will be synthetic gene networks. A gene provides DNA-encoded information for the synthesis of proteins, which cooperate to perform a biological function. With the help of a standardized database of parts like genes, proteins and devices (The BioBrick Fundation website: `http://bbf.openwetware.org/`), it is possible to rationally design circuits made of genes "wired" to one another thanks to regulatory proteins (transcription factors). Such proteins are able to bind DNA on a specific location (e.g. a promoter) and to control (activate or inhibit) the expression of cognate gene(s). Typical circuits are composed of regulated genes coding for a regulating protein. It is therefore possible to design elaborated biological systems performing logic tasks by assembling such circuits. Up to now, most of logic gates and basic combinatorial functions have already been realized with genetic networks [7]. However, combinatorial Boolean gates do not cover all the digital functions. For instance, sequential systems require memories, whose design is a trickier challenge. The development of adapted design tools has therefore become necessary in the field of gene circuits design.

### 1.1 The need for multicellular systems

Synthetic gene networks suffer from two main drawbacks that may limit the complexity of the achievable artificial functions. Firstly, the number of artificial genes that can be added to a given microorganism is generally quite small (some units). Secondly, artificial genes designed for the application should be independent from the genome of the host cell, i.e. any potential interaction (cross- regulation) between the artificial network and the rest of the genome should be minimized. A way to overcome these drawbacks is to split the main function into sub-functions and to implement each one into different host cells [8]. By this means, some components (regulating proteins, promoters ...) may be used several times inside different host cells. However, sub-functions are not completely independent: the signal transfer mechanisms between cells must also be designed [9]. To reach this aim, acyl homoserine lactone (AHL) communication systems [10] of the bacteria *Vibrio fischeri* and *Pseudomonas aeruginosa* are often used [11], as AHL is a molecule small enough to diffuse through the cell plasma membrane and is able to activate gene expression with the help of a cognate protein. If intercellular exchanges are not systematically taken into account in modeling a multi-cellular system yet, a focus should be put on such models since our team showed that the concentration of the signaling molecule (AHL) is crucially dependent on such exchanges and the spatiotemporal behavior of the system.

### 1.2   The need for spatiotemporal simulators

Moreover, synthetic systems are starting to appear that display specific spatiotemporal behavior [12, 13], which have to be taken into account in the modeling. Space can be taken into account in biological models and simulators by different means [14]. Three main approaches exist: particle-centered modeling, space discretization and compartmental modeling. In particle-centered modeling, each instance of each chemical species is modeled as an entity with a given position in a continuous space and a given displacement vector. At every time step, the position of the particle is updated and a new displacement vector is computed either by a draw or by a deterministic equation which may depend on physico-chemical properties. When two particles are close enough, an interaction (binding, degradation, synthesis of another particle) may occur. HSim is an example of simulator that uses such an approach [15]. The complexity of such algorithm grows linearly with the number of chemical species but exponentially with the number of potential interactions. As a consequence, particle-centered modeling can be very useful for the study of elementary mechanisms but it is not adapted to predict the behavior of complex circuits with a high number of particles.

An alternative to particle-centered modeling is to discretize space into a lattice of connected nodes. The concentration of chemical species is computed at every node and every time step as a function of the concentration at neighbor nodes. Many variants of this approach exist. The lattice can be regular or adaptive (mesh size depends on the context), quantities can be Boolean (presence or absence of chemical species in a mesh), discrete (number of chemical particles in a mesh) or continuous values (concentration of chemical species over a mesh), fluxes of particles between nodes can be deterministic or stochastic using partial differential equation (PDE) [16] or cellular automata (CA) [17]. This approach is an equivalent to a T-CAD simulation (e.g. Silvaco) used for the design of semi- conductor devices. Alternatively, COMSOL, which is one of the most used generic PDE solver, has a module dedicated for the modeling of such chemical reactions. Assets and drawbacks of such a modeling approach are well known: a high accuracy versus a low computation time.

The third alternative to model space-dependent biochemical mechanisms is to split space into compartments. In each compartment, the problem is reduced to a space-independent model leading to the use of ordinary differential equations (ODE). The displacement of particles is modeled through unique diffusion equations from compartments to compartments. Compartments are not located in space and the distance and properties of the inter-compartment medium are integrated as fixed parameters in the diffusion model. By this

means, the space- and time-dependent problem is reduced to a time only dependent problem. Models are simplified with regards to the computation time but the price to pay is a decrease of modeling accuracy and the loss of spatial localization of the particle (the concentration of each species is computed only in each compartment but not at every point of the space). Most of the biological simulators, as for instance Virtual Cell [18] use this approach.

In this paper, a mixed modeling approach is described. The main idea consists in using a mesher that divides the space into compartments to enable a compartmental approach (in order to have only time-dependent differential equations). Our mesher is based on two previous studies: a micro-electronics tool used for simulating electro-thermal behavior of integrated circuits [19] and a formalism developed to describe and simulate gene regulatory networks with micro-electronics tools [20]. Those two points are discussed in the first two parts of the present article. Then, the simulation environment is described. Finally, simulation results are given in the last part of the paper.

## 2   Analogy between Biology and Electronics

At the level of a single gene or protein, an analogy can be drawn between micro-electronics and biology [21]. As a gene can lead to the production of a protein, it can be seen as a source of protein and protein synthesis can be represented by a positive current source. Similarly, protein degradation can be associated to a grounded resistance. Protein consumption (by complexation for example) would be represented by a negative current source. According to this analogy, a node in an electronic circuit corresponds to a protein whose concentration is calculated as an equilibrium between the electronic components linked to this node. If the behavior of a cell can be modeled by an electronic circuit, the motion of molecules between cells does not have direct equivalence in electronics. However, this kind of diffusion can be compared to the diffusion of heat in a micro-electronic chip, an issue already tackled in the past.

## 3   Electro-thermal Simulation

The electro-thermal simulation of an integrated circuit becomes unavoidable especially for high density integrated circuits, mixed circuits with integrated power devices or the next generation of 3D integrated circuits. Such a simulation requires two modeling layers: temperature-dependent electronic models for the devices integrated in the circuit on the one hand and a thermal model of the circuit itself (including heat sources near heating devices, heat diffusion inside the silicon, heat accumulation by the chip and heat transfer with the package) on the other.

| Biological behavior | Electronics device |
|---|---|
| Local molecule concentration | Voltage |
| Constant concentration for a given molecule | Voltage source |
| Flux of molecules | Current |
| Constant synthesis or consumption of a molecule | Current source |
| Synthesis or consumption which depends on another molecule | Voltage-driven current source |
| Molecule decay | Resistor |
| Local accumulation of molecules | Capacitor |

**Table 1**: Analogy between Electronics and Biology

There are two main ways to tackle this problem: direct coupling or relaxation [22].

In relaxation approach, a first electrical simulation is carried out with all the devices at room temperature. Then the power dissipated by each device is estimated and a 2D/3D thermal map of the chip is computed with an outsider tool. The mean temperature of each device is then computed from this map and fed back to the electrical simulator. This simulation loop is performed iteratively until convergence is obtained. This approach is straightforward but requires two separated tools. Moreover, very fast changes cannot be considered without the simulation becoming highly time consuming. An alternative consists in a direct coupling between the electrical models of the devices and heat diffusion inside the chip in the same simulator. This can be done by modeling the heat diffusion inside the chip by an equivalent RC network. By this way the complete electro- thermal model of the chip can be simulated with SPICE. The model consists in two coupled sub-circuit, one corresponding to the electrical model of the device (which depends on the temperature) and the other corresponding to the thermal equivalent one (which depends on the power consumption of each device). The spatiotemporal biological simulator presented in this part is based on this approach (Fig. 1).

## 4  Our Spatiotemporal Simulator for Biology

The designed mesher is composed of a suite of modules written in different languages (namely C++, python, Verilog-A and SPICE). The complete workflow is depicted in Fig. 1 and detailed in the following subsections.

### 4.1   Mesher

Firstly, the user has to give a description of the molecule that will diffuse in the input file (see the grey box in Fig. 1). General parameters concerning the lattice (its total size, the maximal and minimal size of a mesh) must be provided, as well as information concerning the biological elements (cells) influencing the concentration of the diffusing molecule (acting as a source or a consummator of the said molecule): each element is characterized by its position, its flux/reaction altering the concentration of molecule and its influence zone (size parameters). The input file is then read by the core module, namely the mesher, a C++ algorithm that will create a list of meshes and their associated nodes, as well as a list of each node's coordinates. To be simulated by the electronic simulator (namely Spectre), these lists need to be translated to a .cir file, which is performed by a Python module. Spectre also requires the model of an elementary mesh, which is described in a Verilog-A file, as well as the location and value of each altering flux of the diffusing molecule, indicated in the input file.

In a 3D space, the algorithm divides first the whole lattice into cubes of the maximal size. As it is an adaptive mesher, it will then additionally divide each cube in contact with the influence zone of a cell into 8 sub-cubes, until only cubes of the desired size are in contact with the influence zone. For each division, the refinement number n is incremented. In order not to have important size differences between neighbor meshes, the algorithm checks whether two neighboring meshes have a n difference strictly superior to 1 and subsequently divides the bigger mesh. The simulation generates a list of each node's value over the total period of simulation. These values can be visualized using a Python module.

### 4.2   .cir generator

The .cir file is the netlist that is supported by SPICE simulator (i.e. Spectre in our case). This file should contain the following elements: i) the definition of the global parameters of the model; ii) The instantiations of elementary mesh; iii) initial concentration map; iv) the boundary conditions; v) instantiation of the biological models of involved cells and vi) simulation command.

The instantiations of elementary mesh is computed directly from the netlist delivered by the mesher. The Verlilog-A model of the elementary mesh is encapsulated in a SPICE subcircuit called `Mesh_unit`. A `Mesh_unit` is instantiated for each line in the netlist file. The k-th mesh is labelled Mk and a list of the 27 connections (26 nodes of the elementary mesh and the reference node which is always 0) is provided (Fig. 2). Unconnected nodes are grounded.
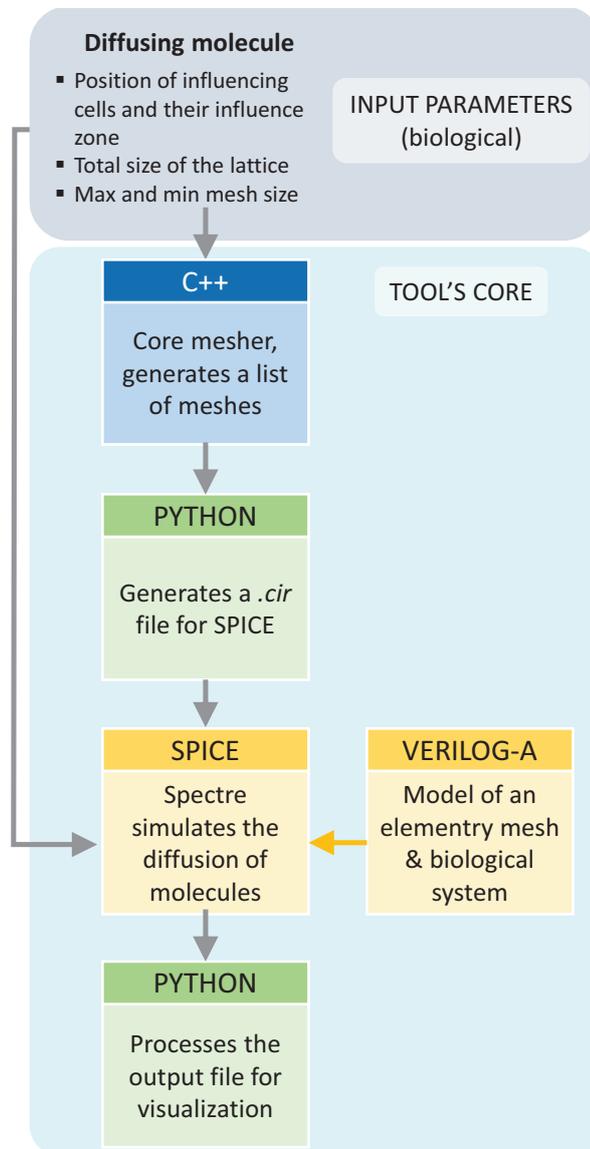
**Figure 1**: The complete workflow of our simulator

Finally, the parameters that are not equal to their default value are specified: ID which is a unique name of the instance (parameter required in order to locate the node inside the Verilog-A model), the degree of refinement n which is equal to 1 by default, 12 parameters XAB which indicate whether the point in

the middle of the edge between corners $A$ and $B$ is connected and 6 parameters XF which indicated that the point at the center of the face $F$ is connected (all the X parameters are set to 0 by default).

By default, the meshed space is not considered as a "closed box" but a part of an infinite space. As a consequence, species continue to diffuse beyond the lattice borders. Thus, the default boundary condition consists in connecting each node on the lattice border to a resistance that models this diffusion and degradation outside the lattice. This resistance is a parameter of the model and its value is shared between each boundary node according to border surface considerations.

The instantiation of cell models is performed according to the content of the input parameter file which contains the position of each cell that may have an influence on the local concentration of the species under consideration. The model of each cell has to be described in an appropriated Verilog-A model beforehand. For each cell, the .cir generator connects one instance of the model of a cell to every node in contact with a cell according to the input parameter file.
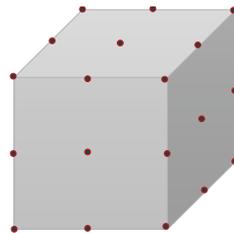


**Figure 2**: Elementary mesh with 26 potential connections: 8 corners (always connected), 12 edge and 6 faces (not connected if the degree of refinement of the neighbour mesh is the same)

### 4.3 Verilog-A model of the elementary mesh

In 3D, an elementary mesh is composed of 26 nodes located at each corner (8), at the middle of each edge (12) and at the center of each face (6). Each node is not necessarily useful and can be "not connected" i.e. connected to the ground (Fig. 2).

Each mesh has 26 parameters. In addition of the 20 parameters described in Sec. IV.B, we define K the capacity, R the decay resistance, D a diffusion coefficient. These three parameters are scaled on the size T of the mesh. $\alpha$ and $\beta$ are two anisotropy coefficients used to compute fluxes inside a mesh.

In 2D, we define fluxes per mesh along the X and the Y axis. On each axis, we define 2 fluxes computed from the difference of concentration (i.e. potential in the electronic model) between the nodes of each side. For instance on axis X the two fluxes $FX1$ and $FX2$ are as follows:

$$FX1 = \frac{(\alpha \cdot C_2 + (1 - \alpha) \cdot C_3) - (\alpha \cdot C_1 + (1 - \alpha) \cdot C_4)}{D} \quad (1)$$

$$FX2 = \frac{(\alpha \cdot C_3 + (1 - \alpha) \cdot C_2) - (\alpha \cdot C_4 + (1 - \alpha) \cdot C_1)}{D} \quad (2)$$

with $C_1$, $C_2$, $C_3$ and $C_4$ being the concentration at the nodes as described on Fig. 3.

These fluxes are distributed on the nodes of each side of the mesh (left and right when considering the X axis and up and down when considering the Y axis). If the node in the middle of the edge of the considered side is not present, the upper left node (4) receives 100% of $FX2$ and the lower left node (1) receives 100% of $FX1$ (as represented on the blue mesh on Fig. 3). On the right side, the lower node receives -100% of $FX1$ and the upper one -100% of $FX2$. If there is a connected node at the middle of the considered edge (e.g. node 41 on the grey mesh of Fig. 3), the upper (4) and lower (1) left nodes only receive 50% of respectively $FX2$ and $FX1$ and the middle node (41) receives the remaining 50% of the two fluxes. In 3D, four fluxes are defined along each axis (X, Y and Z). Fluxes are distributed on the nodes of a face.
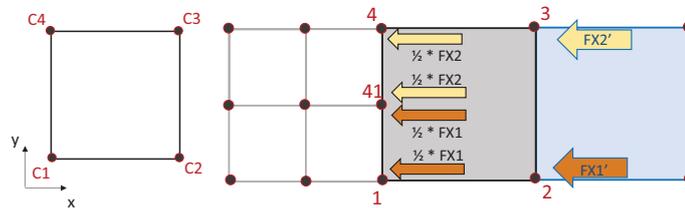


**Figure 3**: Representation of the flux distribution in the 2D case

To each mesh is attributed a capacitance and a resistance related to its size. These values have to be distributed to each node present on the mesh. To this end, we compute the volume attributed to each node, which depends on the position of the node (corner, edge or face) and its neighboring nodes (whether they are present or not). In 2D, the repartition is as follows. When only the corner nodes are present, each receives 1/4 of the total volume (Fig. 4A). If an edge point is added (point 23 in Fig. 4B), the two zones previously attributed to its corner neighbors (blue and red zones) are divided into 2. One half remains attributed to its cognate corner node and the two other halves are attributed to

the edge node, so that the two concerned corner nodes receive 1/8 of the total volume and the edge node receives 1/4. If a second edge node is present on an adjacent edge (node 12 on Fig. 4C), blue zone of node 1 is split into 2, red zone of node 2 as well and the green square (representing 1/16 of the total volume), previously attributed to node 23, is distributed to the two edge nodes. In the end, the edge nodes receive 1/8 + 1/16 + 0.5/16 = 7/32 of the total volume, the corner nodes which only have one adjacent neighbors (point 1 and 3) receive 1/8 and the corner node with 2 adjacent neighbors receives 1/16. When all nodes are present (Fig. 4D), corner nodes receive 1/16 of the total volume and the edge nodes receive 3/16.

This model is extended to 3D space by applying the same rules, except that the volume distribution to a node depends on: i) the number of adjacent edges connected for each corner; ii) the number and the position of adjacent edges and adjacent faces for each edge and iii) the number of adjacent faces for each face.

## 5  Simulation

Once the .cir file and the Verilog-A model have been generated, the circuit can be simulated with every SPICE simulator. In our case, Cadence Spectre simulator is used. Regardless of the kind of simulation performed, it provides ASCII file containing the simulation results. This file is postprocessed by a Python script in order to obtain the graphical results.
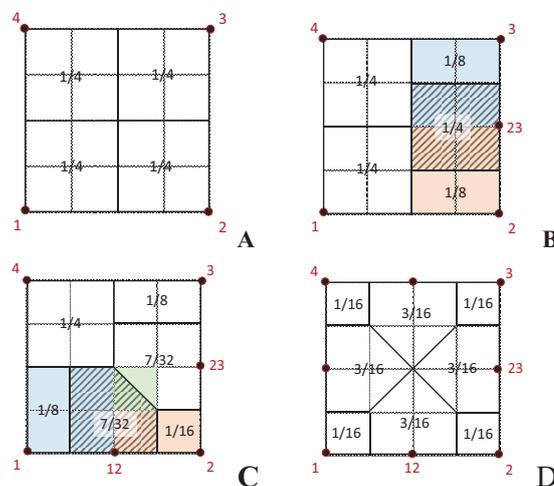


**Figure 4**: Volume distribution inside the 2D mesh as a function pf the number of connected edges: none in sub-figure A, one in B, two adjacent in C and all in D

## 6   Results

To illustrate our tool, we choose to model and simulate a biological band-pass system developed by Basu et al. [12]. This system allows to detect an intermediate concentration of acyl- homoserine lactone (AHL). This system consists in two populations of cells: the senders and the receivers. The senders synthesize and emit isotropically AHL inside a Petri dish. AHL is a small molecule able to diffuse in the gelose of the Petri dish and to enter cells. The receivers react to the concentration of AHL and produce GFP, a green fluorescent protein (the output in this system) if AHL's concentration is comprised between two thresholds. In practice, one or many groups of senders are laid on specific spots on a Petri dish covered with receiver cells.
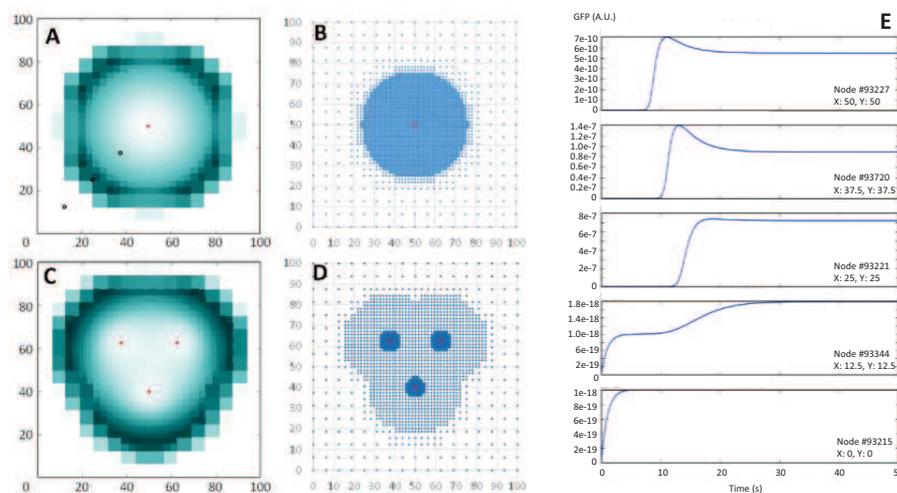


**Figure 5**: Simulation of the Basu system with one (A and B) or three (C and D) sender cells groups. A and C represent the concentration of GFP per mesh (white: no GFP; dark green: maximal concentration of GFP at that time step). B and D represent the lattice configuration: each point is a node of the lattice. Red dot (A and C) or red cross (B and D) represent the sender cell. Blue rings (in A) show the nodes of which the time-dependent GFP concentration is shown on E.

Our tool is used to model the diffusion of AHL in space and time and also to model the synthesis of GFP by the receiver cell depending on the calculated concentration of AHL. At each time step, diffusion is first computed for each mesh in the lattice according to the model discussed above; then for each node of the lattice, GFP expression is calculated according to the Verilog-A biological model of the receiver. Two configurations are tested: one with only one group of sender cells in the middle and another one with 3 groups of sender

cells (Fig. 5). In both cases we use as a base lattice a square of 100x100 units with 16 divisions per axis. In the first case, we added one circular refinement zone with a radius of 25 centered at $x = 50$ and $y = 50$ where the sender cells are located, with a refinement coefficient of 3 (each square in this zone is divided in 4 sub-squares 3 times). Fig. 5B displays all the nodes in the lattice and shows the linear relaxation imposed on the lattice. The concentration of GFP according to space at a given time can be observed in Fig. 5A, with the sender cells represented as a red dot. Time-dependent evolution of GFP synthesis is monitored at 5 nodes: the sender cells (red dot) and nodes in the diagonal toward the lower left corner (blue circles). They are represented in Fig. 5E. It has to be noted that the scale is not the same for each node. We can see that the peak of GFP is propagating from the center toward the borders of the lattice and stops at a certain distance from the center, as expected. For the second configuration, two circular refinement zones were defined per group of sender cells: one with a radius of 3 and a refinement coefficient of 2 and another one with a radius of 20 and a refinement coefficient of 4. They are both centered at their cognate group of senders. Nodes configuration and spatial map of the GFP for this configuration are also given in Fig. 5C and Fig. 5D. Simulation results are in accordance with the results provided by Basu in [12].

For both configurations, the number of equations, of nodes and of required computation time (on a standard computer for a 50 h transient simulation with 182 adaptive time steps) are given in Table 2. In term of computation time, the second configuration requires less CPU time because of its lower number of equations. Indeed, its highly refined zones were smaller compared to the one-group configuration.

|  | 1 sender | 3 senders |
| --- | --- | --- |
| Number of mesh | 3868 | 2 452 |
| Number of receiver models instantiated | 4023 | 2638 |
| Number of nodes in SPICE model | 8046 | 5276 |
| Total number of equation of the SPICE model | 32184 | 21104 |
| CPU time | 54.5 s | 22.5 s |

**Table 2**: Simulation results

## 7 Conclusion and Outlooks

In this paper, a new way to simulate space-dependent biological systems has been presented. This approach is based on an electronic simulator and a very simple adaptive mesher. The results, described in the last section, are in accordance with a case study found in the literature. Compared to existing approaches, our simulator has three main advantages: 1) it is based on a very simple algorithm for the discretization of the space, which facilitates the description of the diffusion phenomena with simple compact models; 2) it provides a direct interface between the diffusion model of molecules and the biological model of each cell, and 3) it uses a SPICE simulation core, which has proven its efficiency for years, especially for systems with a high number of differential equations.

In the future, this tool will allow the simulation and the virtual prototyping of artificial biological systems involving several types of cells that communicate between them through chemical messengers. However, several aspects require further investigations. Firstly, if the generation of the netlist is automated, the configuration of the tool and the integration of third-party biological model is a hand-made process. As a consequence, the software cannot be easily handled by engineers that are not accustomed to SPICE and Verilog-A. The tool has to be hidden behind a biologist-friendly interface or should ideally be able to directly support biological descriptions (in a dedicated language as SBML). Secondly, the issue of validation may arise. The validation of the tool in comparison with experimental results is very tricky as most of the quantities provided by the simulator are not biologically observable: in our case, the spatial distribution of GFP is the only observable parameter. One solution would be to validate our results in comparison with existing finite-element simulation tools. This work is ongoing. Finally, the simulation may become problematic, especially for multi-species problems requiring tightened meshes. The deployment of the tool on GPU is also worth of investigation.

## References

[1] A. S. Khalil and J. J. Collins, "Synthetic biology: applications come of age.," *Nat. Rev. Genet.*, vol. 11, no. 5, pp. 367–79, May 2010.

[2] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli.," *Nature*, vol. 403, no. 6767, pp. 339–42, Jan. 2000.

[3] M. A. Fisher, K. L. McKinley, L. H. Bradley, S. R. Viola, and M. H. Hecht, "De novo designed proteins from a library of artificial sequences

function in Escherichia coli and enable cell growth.," *PLoS One*, vol. 6, no. 1, p. e15364, Jan. 2011.

[4] J. Nielsen and J. D. Keasling, "Synergies between synthetic biology and metabolic engineering.," *Nat. Biotechnol.*, vol. 29, no. 8, pp. 693–5, Aug. 2011.

[5] V. Noireaux, Y. T. Maeda, and A. Libchaber, "Development of an artificial cell, from self-organization to computation and self-reproduction," *Proc. Natl. Acad. Sci.*, vol. 108, no. 9, pp. 3473–3480, Feb. 2011.

[6] D. P. Ho, H. H. Ngo, and W. Guo, "A mini review on renewable sources for biofuel," *Bioresour. Technol.*, vol. 169, pp. 742–749, Jul. 2014.

[7] S. Ausländer, D. Ausländer, M. Müller, M. Wieland, and M. Fussenegger, "Programmable single-cell mammalian biocomputers.," *Nature*, vol. 487, no. 7405, pp. 123–7, Jul. 2012.

[8] K. Brenner, L. You, and F. H. Arnold, "Engineering microbial consortia : a new frontier in synthetic biology," no. July, pp. 483–489, 2008.

[9] S. Regot, J. Macia, N. Conde, K. Furukawa, J. Kjellén, T. Peeters, S. Hohmann, E. de Nadal, F. Posas, and R. Solé, "Distributed biological computation with multicellular engineered networks.," *Nature*, vol. 469, no. 7329, pp. 207–11, Jan. 2011.

[10] K. M. Gray, L. Passador, B. H. Iglewski, and E. P. Greenberg, "Interchangeability and specificity of components from the quorum-sensing regulatory systems of Vibrio fischeri and Pseudomonas aeruginosa.," *J. Bacteriol.*, vol. 176, no. 10, pp. 3076–80, May 1994.

[11] H. Song, J. Ozaki, C. H. Collins, M. Barnet, F. H. Arnold, and S. R. Quake, "REPORT A synthetic Escherichia coli predator – prey ecosystem," no. 187, pp. 1–8, 2008.

[12] S. Basu, Y. Gerchman, C. Collins, F. Arnold, and R. Weiss, "A synthetic multicellular system for programmed pattern formation," *Nature*, vol. 434, no. April, 2005.

[13] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires,'" *Nature*, vol. 469, no. 7329, pp. 212–215, 2011.

[14] K. Takahashi, S. N. V. Arjunan, and M. Tomita, "Space in systems biology of signaling pathways–towards intracellular molecular crowding in silico.," *FEBS Lett.*, vol. 579, no. 8, pp. 1783–8, Mar. 2005.

[15] P. Amar, G. Bernot, and V. Norris, "HSIM: a simulation programme to study large assemblies of proteins," *J. Biol. Phys. Chem.*, vol. 4, pp. 79–84, 2004.

[16] J. Schaff, C. C. Fink, B. Slepchenko, J. H. Carson, and L. M. Loew, "A general computational framework for modeling cellular structure and function.," *Biophys. J.*, vol. 73, no. 3, pp. 1135–46, Sep. 1997.

[17] M. Madec, Y. Gendrault, C. Lallement, and J. Haiech, "A game-of-life like simulator for design-oriented modeling of BioBricks in synthetic biology.," *Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2012, pp. 5462–5, Jan. 2012.

[18] L. M. Loew and J. C. Schaff, "The Virtual Cell: a software environment for computational cell biology.," *Trends Biotechnol.*, vol. 19, no. 10, pp. 401–6, Oct. 2001.

[19] J.-C. Krencker, J. Kammerer, Y. Hervé, and L. Hébrard, "Direct electro-thermal simulation of integrated circuits using standard CAD tools," pp. 1–4, 2010.

[20] Y. Gendrault, M. Madec, C. Lallement, and J. Haiech, "Modeling biology with HDL languages: A first step toward a genetic design automation tool inspired from microelectronics," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1231–1240, 2014.

[21] Y. Gendrault, M. Madec, C. Lallement, F. Pecheux, and J. Haiech, "Synthetic biology methodology and model refinement based on micro-electronic modeling tools and languages.," *Biotechnol. J.*, vol. 6, no. 7, pp. 796–806, Jul. 2011.

[22] G. Digele, S. Lindenkreuz, and E. Kasper, "Fully coupled Dynamic Electro-thermal Simulation," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 5, no. 3, pp. 250–257, 1997.

# Computational algorithms as biological switches

Hernansaiz-Ballesteros, R.D.[1], Dalchau, N.[2], Cardelli, L.[2,3],
Csikász-Nagy, A.[1,4]

[1] Randall Division of Cell and Molecular Biophysics and Institute of Mathematical
and Molecular Biomedicine, King's College London, London, SE1 1UL, UK
[2] Microsoft Research, 21 Station Road, Cambridge CB1 2FB, UK
[3] Department of Computer Science, University of Oxford, Wolfson Building,
Parks Road, Oxford OX1 3QD, UK
[4] Pázmány Péter Catholic University, Faculty of Information Technology and Bionics,
Budapest, Hungary

## *Abstract*

Biological systems have developed a wide range of mechanisms to respond to changes in their environment. Biological switches are mechanisms which drive a change in the functional state of a system in an all-or-none fashion. Biological switches produce a reliable and robust transition between states, sometimes generating an irreversible transition.

Current eukaryotic switches usually contain several components with multiple positive feedback loops. This level of complexity could have been reached by an evolutionary process from a simple system. The simplest possible system could have been a single molecule that regulates itself and it can go through different functional states. Most of the quickly responding switches work on the post- translational level. Phosphorylation has been proposed as an ancient post-translational mechanism, which could have defined phosphorylated and dephosphorylated states.

It was previously shown that simple systems based on single autocatalytic element could behave like a switch. Therefore, maybe there exist an evolutionary way of obtaining complex networks from simpler ones. We investigate from a computational perspective, the increasing complexity from simple systems to complex ones and how their key properties could have been kept through evolution.

## 1 Introduction

Biological systems have developed signal processing mechanisms to measure and respond to changes in their environment. To certain stimuli, feedback mechanisms trigger a readjustment of the stimulated molecule, either directly or through a number of connected elements. Thus, environmental information

is used to modify the state of downstream pathways, adapting the system to generate an appropriate response.

### 1.1 Feedback loops as building blocks of regulatory networks

Feedback loops (FBL) constitute a basic relationship among molecules to construct complex behaviours [1-7]. There are two classes of FBL, positive or negative, depending on the number of negative interactions (inhibitions) [1, 2].

Negative feedback loops (NFBL) contain an odd number of negative interactions. These systems counteract the effect of stimuli, sometimes working like a biological thermostat (homeostasis) [1-5,8]. When there is a long enough delay in the feedback pathway, NFBL can also create oscillations [1-5,8,9]. Positive feedback loops (PFBL) contain an even number of negative interactions and/or positive interactions only. These systems are forced to choose between the possible states, and may create switches [1-5, 10]. Biological switches are able to convert graded inputs into on/off responses, when the input reaches a certain threshold [1-5, 11-14].

Biological switches present interesting features to take advantage of. *Reliability* and *robustness* are key properties. These systems cannot be arrested in undecided states, they should choose between the possible states and ensure the selected state will be reached despite small perturbations [15]. Thus, hysteretic behaviour and bistability are generated: once a state is reached, the system "remembers" the previous state, and only major changes in the input can return it to the other state [16]. A related feature is the *speed* in the transition between states. The transition should be sufficiently fast to reduce the time the system is undecided. At the same time, the system should remain in a given state for a sufficient amount of time to perform any necessary functions.

Reliability, robustness and speed are features also attractive for designing computational algorithms that look to perform efficient all-or-none decisions [17-19]. Recently it was shown how a computational algorithm (the Approximate Majority) presents similar dynamical features to a key transition of the cell cycle regulatory network [20].

### 1.2 The cell cycle and the approximate majority algorithm

The Approximate Majority computational algorithm (AM) is a fast and simple population protocol of distributed computing [21]. Given an initial population of elements in either of two decided (or active) states ($x_0$ or $x_2$), AM describes how to drive the initial population into a final one where all elements are in the same state (Figure 1, Figure 2A,C). The key piece of this algorithm is the introduction of a third state, the *blank* or *undecided* state ($x_1$). From the

undecided state both decided states can be reached, attending to a set of rules. If the undecided state meets any of the decided states, it takes the state of the decided state $(x_0 + x_1 \rightarrow x_0 + x_0; x_2 + x_1 \rightarrow x_2 + x_2)$; if opposite states meet each other, they both become undecided $(x_0 + x_2 \rightarrow x_1 + x_1)$ [21]. This system shows three positive feedback loops (two pure positive and one double negative); each of the decided states activates themselves and prevent the growth of the other state (Figure 1).
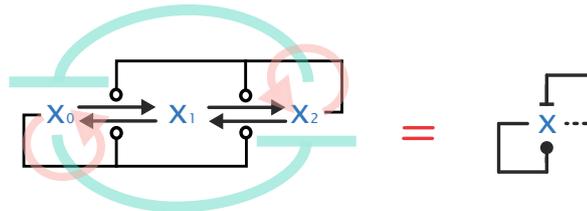


**Figure 1**: **Approximate majority.** Extended (left) and condensed (right) versions of the AM. $x_0$ activates itself and $x_2$ activates itself, but at the same time inhibits the other form. The extended wiring diagram can be collapsed into the condensed version. The dashed line of the condensed wiring diagram indicates the action of $x_2$, which produces an inhibition (bar-end line) of the $x_0$ state. The solid line indicates the action of $x_0$, which produces an activation (ball-end line) of the $x_0$ state.

The eukaryotic cell cycle is divided into several phases (G1, S, G2, M). The transitions between phases are regulated by complex networks of kinases and phosphatases, generating positive feedback loops [22, 23]. The transition from late interphase (G2 phase) to Mitosis (M phase) is termed as G2/M transition, and is driven by Cdk-cyclin complexes (Cdk). At this transition, Cdk is regulated by two positive feedback loops. Cdk is in a low activation state before the transition. It is slightly activated by its phosphatase activator, Cdc25, and mainly inactivated by its kinase inhibitor, Wee1. Once Cdk reaches a critical, but still low level of activation, the PFBL flip to a state in which the level of activated Cdk and Cdc25 are high, and Wee1 goes down [23-28]. This state will be kept until the cell finishes mitosis [23]. The main regulatory network (Cdk-Wee1-Cdc25) is extended by the presence of the Greatwall kinase and the phosphatases PP2A and PP1 [29, 30]. PP2A/PP1 indirectly inhibits Cdk by inactivation of its activator (Cdc25) and activation of its inhibitor (Wee1). Thus, they set the critical threshold for the flip of the switch. At the same time, Cdk indirectly inhibits PP2A/PP1 through Greatwall (Figure 2).

The AM algorithm behaves like a reliable, robust and fast switch. This behaviour has the same characteristic features as biological switches. Moreover, AM, with only one autocatalytic molecule, produces similar dynamical properties to complex systems like the G2/M transition [20, 31] (Figure 2).

Thus, it could be expected that early biological switches looked more like the AM system. This system has subsequently evolved to increase complexity to reach the G2/M module. We are therefore investigating the possibility of an evolutionary pathway between AM and more complicated biological networks.
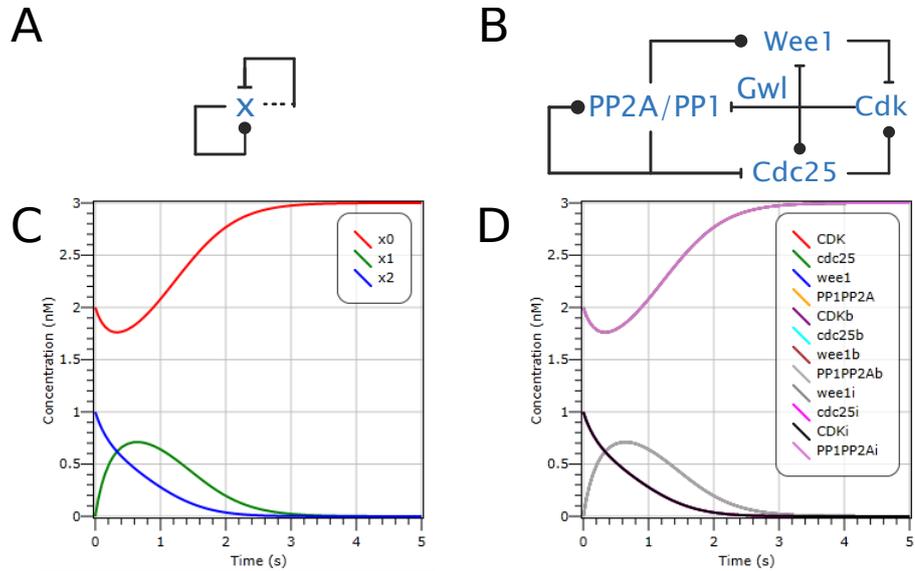


**Figure 2**: **Comparison of the AM and an extended cell cycle model**. **A, B**: Model of AM and the G2/M transition extended with Gwl and PP2A/PP1, **C, D**: Time-course simulation of the models on A and B, respectively. Each node in the wiring diagrams (A,B) is an influence node (see Methods below), so the time-course diagram show for each node three species.

## 2   Methods and materials

The presented models are based on the principles of influence networks [4, 31-33]. Influence networks are abstractions of more detailed biological interactions; they capture only the effects between species [32]. These effects can be of any type, with any biological meaning, but usually they are simplified to activation and inhibition relationships [32].

Influence networks are represented as a graph of *influence nodes* (species) and *influence edges* (reactions). Each influence node can interact in at most four ways with another influence node (Figure 3). Two interactions influence another node (output), and the other two represent the actions on the node (inputs). Influence edges connect one input (In0 and In2) with one output (Out0 and Out2).
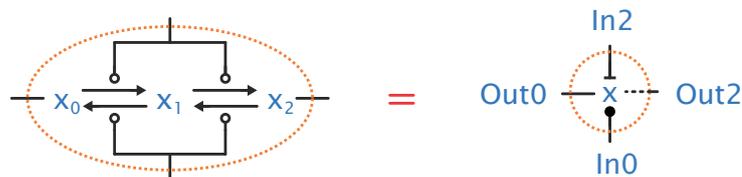
**Figure 3**: **Influence network notation**. The triple motif (left) is composed by three species ($x_0$, $x_1$, $x_2$) connected by four reactions. Only the decided states ($x_0$, $x_2$) can generate an action over another species. These actions catalyse the reactions to generate decided states (upper and bottom interaction in the left diagram). The triple motif is compressed in an influence node (right). Each influence node (x) has four ways of communication, two outputs (Out0, Out2) and two inputs (In0, In2). The solid line (Out0) indicates the action that $x_0$ applies over a node; the dashed line (Out2) shows the action of $x_2$ over a node; the ball-end line indicates the action received from a "0" species; the bar-end line indicates the action received from a "2" species.

Influence nodes are modelled as a triplet motif: three chemical species ($x_0$, $x_1$, $x_2$) connected by four reactions (Figure 3). As in the description of AM (Figure 1), $x_0$ and $x_2$ are the decided or active states; they push other influence nodes towards their decided states. The intermediate step ($x_1$) represents the undecided or blank state, and it introduces nonlinearity in the system. The undecided state is connected just with the decided states, and it does not interact in any other way with the rest of the network [32]. In biological systems, this intermediary state can represent the multisite modifications of proteins, leading to their functional states [34, 35].

Influence networks are interpreted as finite sets of irreversible reactions over a finite sets of species [32]. As example, the conversion of $x_0$ in $x_1$, which is catalysed by $x_2$, is represented by the reaction $x_0 + x_2 \rightarrow x_2 + x_1$ (Figure 1). The reaction rates are constants and mass action kinetics are applied to solve these equations for each triplet motif (influence node). Solving these mass action equations for each influence node at steady yields a generalized Hill function of coefficient 2 [32].

Each model is expressed through a set of mass action equations. They are written in Language for Biochemical Systems (LBS) [36], a programming language used to generate reaction models and afford molecular descriptions. This language is implemented into Visual GEC [36, 37]. Visual GEC (`http://research.microsoft.com/gec/`) is an easy, user-friendly interface tool developed by Microsoft Research for the design and simulation of biological circuits or devices.

All the simulations have been running using equal rates ($k = 1$) for all the re-actions, highlighting that the qualitative behaviour of the systems is dependent on the wiring of the network, but not on actual rates. The Chemical Master Equation (CME) approach of Visual GEC has been used to analyse the *state* of chemical systems [38, 39]. The Chemical Master Equation is an equation that determines, for each species, the probability of having a specific molecular population at a given future time [40].

## 3   *Results*

Phosphorylation has been proposed as the first post-translational modification in proteins [41, 42]. From a biological point of view, AM could be thought as an autocatalytic system which resembles phosphorylation/dephosphorylation events. One active state phosphorylates when it is phosphorylated (kinase activity), and the other active state dephosphorylates when it is dephosphorylated (phosphatase activity) (Figure 1). This AM behaves like a bifunctional enzyme that can work both as a kinase and a phosphatase. Such molecules are wide spread in prokaryotes [43].

To keep the same nomenclature, the active state "0" would be the catalytic state that phosphorylates (kinase), and the active state "2" would be the one that dephosphorylates (phosphatase). The ball-end edge indicates phosphorylation, and the bar-end edge represents dephosphorylation (Figure 4).

Setting AM as a possible ancient biological switch (Figure 4A), an event of duplication (duplicate a node and its edges) and some loss of function (remove edges) are applied to reach a more complex network (Figure 4). By a duplication event, the autocatalytic element ($x$) is amplified (Figure 4B). The new element ($z$) has the same interactions as the initial one, but also both elements are connected, as they are exactly the same. Thus, $x$ and $z$ phosphorylate ($x_0$ and $z_0$) and dephosphorylate ($x_2$ and $z_2$) both themselves and each other.

After the duplication event, events of loss of function reduce the number of interactions between the elements (Figure 4CD). This diminution leads the system to maintain just one active state per element, so they either phosphorylate or dephosphorylate. Therefore, instead of having two catalytic states per element, there will be just one catalytic state per element, which is more likely to be found in natural systems. The new system resembles a mutual inhibition network (MI) [3, 32]. In MI, phosphorylation by $x_0$ drives the system to the $x_0$ and $z_0$ active states (phosphorylated forms of $x$ and $z$). Thus, $x$ is active because $x_0$ has catalytic activity, whereas $z_0$ is inhibited (this state does not have catalytic activity).
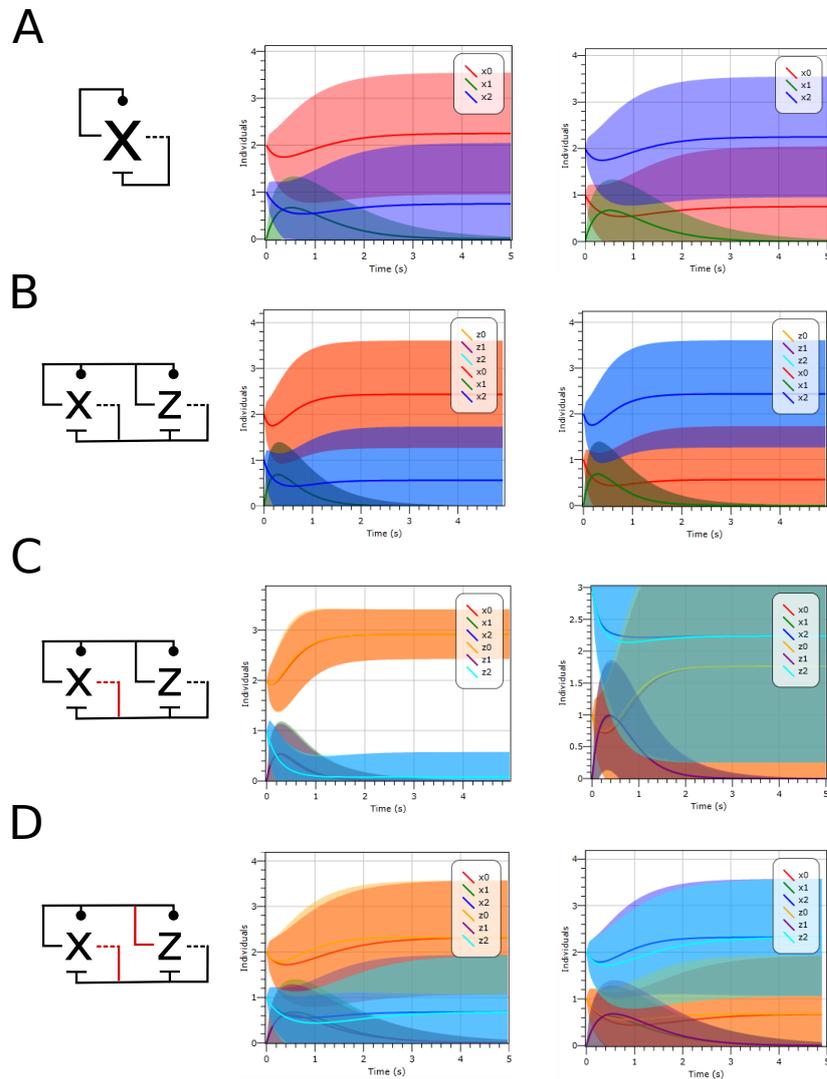
**Figure 4**: **Evolution from AM to MI**. Each row (A, B, C, D) includes the wiring diagram and two time-course diagrams to show the bistability of the system; the first time-course diagram exhibits the dynamics when the state that phosphorylates wins ($x_0$, $z_0$), and the second time-course diagram shows the dynamics when the state that dephosphorylates wins ($x_2$, $z_2$). The red edges indicate the affected interactions by the loss of function events. Ribbons indicate the variance per each trace. **A**: Approximate Majority (AM); **B**: duplicated AM; **C**: loss of dephosphorylation activity in one molecule; **D**: Mutual Inhibition.

Conversely, dephosphorylation by $z_2$ activates $z$ ($z_2$ moves the system to $z_2$, which holds catalytic activity) and inhibits $x$ ($z_2$ moves the system to $x_2$, which does not have catalytic activity). Thus, a system of an antagonistic kinase-phosphatase pair could have evolved from a bifunctional autocatalytic molecule.

All networks from the initial AM to the MI system show bistability, and the initial conditions determine which steady state will be reached. As the complexity increases, the dynamics are similar but the noise (variance) changes [39]. When AM is duplicated, the noise is reduced. The loss of interactions in the intermediary system (Figure 4C) destabilizes it, so one steady state is more robust than the other (Figure 4C). When the second edge is removed, the system is symmetric again (Figure 4D), and the original AM behaviour is restored.

## 4  Discussion

We have observed that AM-like systems can arise from direct autocatalytic processes of two differentiated states of a single molecule. Autocatalytic processes have been proposed a basis for developing complex interactions [44-46]. Also, autocatalytic systems are likely to appear by chance [45]; they can maintain themselves inside a reaction soup of molecules, and stabilize themselves in time [45].

One of the key parameters suggested for the evolution of prebiotic networks is information control [5, 46]. Phosphorylation is one of the oldest post-translational mechanisms [41, 42]. As AM can represent phosphorylation/dephosphorylation events, one external signal that could have driven the system is the level of environmental phosphate. Current kinase's activity depends on the ATP/ADP level [47, 48]. If the ratio is high, kinases use ATP to phosphorylate their substrates; if the ratio is low, kinases catalyse the reverse reaction. Thus, a unique molecule, like the proposed AM, can phosphorylate (acting as a kinase) or dephosphorylate (acting as phosphatase), coupling with the function of measuring the phosphate level. In an environment of high phosphate, the element will remain phosphorylated, but if there is an absence of phosphate, the element will be mainly found as dephosphorylated. AM acting as a switch could have served as a critical ATP/ADP sensor, switching only at critical thresholds.

Duplication and mutation events are fundamental in the evolution of biological networks. Duplication events result in a new copy which retains the original function, whereas mutation events can cause the loss of the old function or the gain of a new one. As soon as a new copy appears, one can keep the original function and the other is susceptible to changes, without reducing fitness of cells. However, both copies have the same probability of obtaining mutations [49, 50].

In the proposed models, after the duplication event, the dynamics of the system is kept and the noise is reduced. This therefore increases the stability of the switch. Despite this reduction, the cost of energy needed to keep both copies is higher, and sometimes it might be not worth it. Thus, if reduction in the noise level is not enough to keep both copies, mutational events rapidly occur to silence one of them. However, as both copies could be equally affected, both could accumulate variations. This could lead to a new system, as the proposed MI, which keeps similar dynamical properties than the original AM.

It has been proposed that evolution works through building blocks [4, 48,51,52]. From a core mechanism in which layers of complexity are added (e.g. kinases) [52], to give rise to networks of conserved, molecularly complementary modules (e.g. molecules involved in glucose metabolism) [51]. Eukaryotic kinases have evolved from an initial core [47, 48, 52]; they have been tracked to find the possible ancestor. RIO kinase, an atypical kinase, has been proposed as the ancestor of the canonical eukaryotic protein kinases [53, 54]. RIO kinase, apart from its kinase activity, it presents ATPase activity and the ability of autophosphorylation [53-55].

Our findings, together with these biological evidences, suggest AM as a possible ancient biological switch of just one element in a direct autocatalysis. This system could have evolved to other complex systems by events of duplication and mutations, while always keeping the same dynamical properties of the ancient one. If AM could have been a bifunctional enzyme, duplications and mutations could have separated these functions, obtaining monofunctional enzymes with multiple sites of modifications. Here we have showed a possible way of evolution, but in the biological context several others could have appeared, tested and kept or erased.

## 5 Acknowledgements

## References

[1] Thomas R. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Numerical methods in the study of critical phenomena* 1981, Springer Berlin Heidelberg 180–193

[2] Thomas, R., & d'Ari, R: Biological feedback. *CRC press*, 1990.

[3] Tyson JJ, Chen KC, Novak, B: Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current opinion in cell biology* 2003, 15(2):221–231.

[4] Tyson JJ, Novák B: Functional motifs in biochemical reaction networks. *Annual review of physical chemistry* 2010, 61:219.

[5] Thomas, R: Laws for the dynamics of regulatory networks. *International Journal of Developmental Biology* 1998, 42, 479–485.

[6] Soulé C. Graphic requirements for multistationarity. *ComPlexUs* 2003,1(3):123–133.

[7] Bernot G, Comet JP, Richard A., Guespin, J. Application of formal methods to biological regulatory networks: extending Thomas' asynchronous logical approach with temporal logic. *Journal of theoretical biology* 2004, 229(3):339–347.

[8] Griffith JS. Mathematics of cellular control processes I. Negative feedback to one gene. *Journal of theoretical biology* 1968, 20(2), 202–208.

[9] Novák B, Tyson JJ: Design principles of biochemical oscillators. *Nature reviews Molecular cell biology* 2008, 9(12), 981–991.

[10] Griffith JS. Mathematics of cellular control processes II. Positive feedback to one gene. *Journal of Theoretical Biology* 1968, 20(2), 209–216.

[11] Brandman O, Meyer T: Feedback Loops Shape Cellular Signaling in Space and Time. *Science* 2008, 322(5900):390–395.

[12] Brandman O, Ferrell JE, Li R, Meyer T: Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science* 2005, 310(5747):496–498.

[13] Ferrell JE: Feedback regulation of opposing enzymes generates robust, all-or-none bistable responses. *Current Biology* 2008, 18(6):R244–R245.

[14] Xiong W, Ferrell JE: A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature* 2003, 426(6965):460–465.

[15] Kitano H: Biological robustness. *Nature Reviews Genetics* 2004, 5(11):826–837.

[16] Angeli D, Ferrell JE, Sontag ED: Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101(7):1822–1827.

[17] Kitano H: Computational systems biology. *Nature* 2002, 420(6912):206–210

[18] Priami C: Algorithmic systems biology. *Communications of the ACM* 2009, 52(5), 80–88.

[19] Navlakha S, Bar-Joseph Z: Algorithms in nature: the convergence of systems biology and computational thinking. *Molecular systems biology* 2011, 7(1):546.

[20] Cardelli L, Csikász-Nagy A: The cell cycle switch computes approximate majority. *Scientific reports* 2012, 2.

[21] Angluin D, Aspnes J, Eisenstat D: A simple population protocol for fast robust approximate majority. *Distributed Computing* 2008, 21(2):87–102.

[22] Romanel A, Jensen LJ, Cardelli L, Csikász-Nagy A. Transcriptional regulation is a major controller of cell cycle transition dynamics. *PloS one* 2012, 7(1):e29716.

[23] Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A: Irreversible cell-cycle transitions are due to systems-level feedback. *Nature Cell Biology* 2007, 9(7):724–728.

[24] Ferrell JE, Ha SH: Ultrasensitivity part III: cascades, bistable switches, and oscillators. *Trends in biochemical sciences* 2014, 39(12):612–618.

[25] Kim SY, Ferrell JE: Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. *Cell* 2007, 128(6):1133–1145.

[26] Morgan DO: Principles of CDK regulation. *Nature* 1995, 374(6518), 131–134.

[27] O'Farrell PH: Triggering the all-or-nothing switch into mitosis. *Trends in cell biology* 2001, 11(12):512–519.

[28] Trunnell NB, Poon AC, Kim SY, Ferrell JE: Ultrasensitivity in the regulation of Cdc25 by Cdk1. *Molecular cell* 2011, 41(3):263–274.

[29] Lorca T, Castro A: The Greatwall kinase: a new pathway in the control of the cell cycle. *Oncogene* 2013, 32(5):537–543.

[30] Mochida S, Maslen SL, Skehel M, Hunt T: Greatwall phosphorylates an inhibitor of protein phosphatase 2A that is essential for mitosis. *Science* 2010, 330(6011):1670–1673.

[31] Fages F, Soliman S: From reaction models to influence graphs and back: a theorem. *Formal Methods Syst Biol* 2008, LNCS 5054:90–102.

ADVANCES IN SYSTEMS AND SYNTHETIC BIOLOGY

[32] Cardelli L: Morphisms of reaction networks that couple structure to function. *BMC systems biology* 2014, 8(1):84.

[33] Verdugo A, Vinod PK, Tyson JJ, Novak B: Molecular mechanisms creating bistable switches at cell cycle transitions. *Open Biol* 2013, 3(3):120179. 10.1098/rsob.120179

[34] Cohen P. The regulation of protein function by multisite phosphorylation–a 25 year update. *Trends in biochemical sciences* 2000, 25(12), 596–601.

[35] Markevich NI, Hoek JB, Kholodenko BN. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of cell biology* 2004, 164(3), 353–359.

[36] Pedersen M, Plotkin GD: A language for biochemical systems: design and formal specification. *In Transactions on computational systems biology XII* (pp. 77–145), 2010. Springer Berlin Heidelberg.

[37] Pedersen M: Modular Languages for Systems and Synthetic Biology. *Laboratory for Foundations of Computer Science, School of Informatics, University of Edinburgh* 2010, PhD thesis.

[38] Dalchau N, Chandran H, Gopalkrishnan N, Pillips A, Reilf J: Probabilistic analysis of localized DNA hybridization circuits. *ACS Synthetic Biology* 2015, 4, 898913.

[39] Cardelli L, Csikász-Nagy A, Dalchau N, Tribastone M, Tschaikowski M. Noise Reduction in Complex Biological Switches. *Scientific Reports* 2016, 6, 20214; doi: 10.1038/srep20214

[40] Gillespie DT: Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 2007, 58:35–55.

[41] Kennelly PJ: Protein kinases and protein phosphatases in prokaryotes: a genomic perspective. *FEMS microbiology letters* 2002, 206(1):1–8.

[42] Kennell PJ: Protein Ser/Thr/Tyr phosphorylation in the archaea. *Journal of Biological Chemistry* 2014, 289(14):9480–9487.

[43] Porter SL, Roberts, MA, Manning, CS, Armitage, JP. A bifunctional kinase-phosphatase in bacterial chemotaxis. *Proceedings of the National Academy of Sciences* 2008, 105(47):18531–18536.

[44] Plasson R, Brandenburg A, Jullien L, Bersini H: Autocatalysis: At the root of self-replication. *Artificial life* 2011, 17(3):219–236.

[45] Jain S, Krishna S. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proceedings of the National Academy of Sciences* 2001, 98(2):543–547.

[46] Nghe P, Hordijk W, Kauffman, SA, Walker, SI, Schmidt, FJ, Kemble, H, ... Lehman, N. Prebiotic network evolution: six key parameters. *Molecular BioSystems* 2015, 11(12):3206–3217.

[47] Endicott JA, Noble ME, Johnson LN: The structural basis for control of eukaryotic protein kinases. *Annual review of biochemistry* 2012, 81:587–613.

[48] Taylor SS, Kornev AP: Protein kinases: evolution of dynamic regulatory proteins. *Trends in biochemical sciences* 2011, 36(2):65–77.

[49] Behe MJ, Snoke DW: Simulating evolution by gene duplication of protein features that require multiple amino acid residues. *Protein Science* 2004, 13(10):2651–2664.

[50] Veitia RA, Potier MC: Gene dosage imbalances: action, reaction, and models. *Trends in biochemical sciences* 2015, 40(6):309–317.

[51] Root-Bernstein R. A modular hierarchy-based theory of the chemical origins of life based on molecular complementarity. *Accounts of chemical research* 2012, 45(12):2169–2177.

[52] Oruganty K., Kannan N. Design principles underpinning the regulatory diversity of protein kinases. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2012, 367(1602):2529–2539.

[53] LaRonde NA: The ancient microbial RIO kinases. *Journal of Biological Chemistry* 2014, 289(14):9488–9492.

[54] Bahmanjah S, LaRonde N. Investigation of the evolutionary relationship between Rio2 kinases and the canonical eukaryotic protein kinases using x-ray crystallography. *The FASEB Journal* 2015, 29(1 Supplement):573–33.

[55] LaRonde NA. Structure of the eukaryotic atypical RIO kinases reveals ATPase-dependent mechanism in ribosome biogenesis. *The FASEB Journal* 2013, 27(1_MeetingAbstracts):1045–5.

# Investigating the metabolic limitations of Chinese hamster ovary (CHO) cells for recombinant protein production under fixed nutritional conditions

Sarah Galleguillos[1,2], Michael Hanscho[1,2], Nicole Borth[1,2],
Jürgen Zanghellini[1,2]

[1] Austrian Centre of Industrial Biotechnology, Muthgasse 11, A1190 Vienna
Austria

[2] Department of Biotechnology, University of Natural Resources and Life Sciences
Vienna, Austria

## *Abstract*

Most of the biotherapeutic recombinant proteins are currently being produced in Chinese hamster ovary (CHO) cells. The increasing demand of such biopharmaceuticals has driven major efforts in process optimization. Further improvements will only be achievable with a proper systems understanding of CHO cell metabolism. The availability of the genomic sequencing information for CHO amplified and facilitated a community effort to build a genome-scale metabolic model. Here, we used this model and flux-balance analysis to investigate the effect of the amino acid composition of the protein of interest on its maximal yield for a given host.

## 1 Introduction

Mammalian cell culture is of major relevance for the production of vaccines, growth factors, hormones, interferons, enzymes and monoclonal antibodies for therapeutic applications [1]. The majority of these biopharmaceuticals are obtained from fed-batch cultures of Chinese hamster ovary (CHO)-derived cell lines. These hosts have proven to be appropriate expression systems for complex recombinant proteins due to their capability of performing human-compatible post-translational modifications (e.g. glycosylation) and correct folding.

Given the high production needs and elevated costs for many of these compounds [2], increasing efforts have been focused on optimizing the efficiency and yield of the culture process. Different approaches include media optimization [3, 4], feeding strategies and genetic engineering [5, 6]. Although these methods remarkably improved the production yield, there is a lack of systematic analysis tools for CHO cells.

Constraint-based reconstruction and analysis (COBRA) approaches were shown to deliver detailed insight into cellular metabolism of many prokaryotes and eukaryotes, most prominently *E. coli* [7] and *S. cerevisiae* [8]. However, a prerequisite for any COBRA approach is a comprehensive, organism specific in silico reconstruction of the corresponding metabolism, i.e. a genome-scale metabolic model (GSMM). Such a GSMM was missing for CHO and only recently became available [9].

### CHO, a community driven GSMM of CHO

Several research groups joined forces and generated a single, comprehensive CHO GSMM combining currently available knowledge, data and unpublished reconstruction work. In a first step the gene-protein-reaction relationship for a carefully verified reconstruction for *Homo sapiens* [10, 11], that provides the basis for the CHO community reconstruction, was manually curated. The improved annotation was needed to correctly extract CHO homologous reactions from the human reconstruction to build an appropriate scaffold for the CHO GSMM and to be able to compare and merge the research group specific versions. Another round of manual curation followed the merge process. Including CHO specific reactions based on literature data added further knowledge to the model . Finally strain specific model versions were generated using metabolomics and RNA-Seq data.

The latest version of iCHO consists of 2,341 unique metabolites, 6,663 reactions and 1,766 annotated genes. Table 1 shows a comparison with other published GSMMs. The ratio of annotated genes per reaction was found to be similar for the human GSMM and iCHO, but significantly lower than in the GSMMs of *S. cerevisiae* and *E. coli*. To some extent the lower ratio can be explained by the huge number of not annotated transporters that are needed to connect different compartments in mammalian cells. The comparison of the values suggests that significant efforts are needed to improve the number of annotated reactions.

### Objective

In the following we will use iCHO to study the potential of CHO cells for recombinant protein production. Our aim was to investigate for a given cell line the effect of the product's amino acid (AA) composition onto its productivity. It is well known that CHO's productivity is dependent on the recombinant protein. This effect is somewhat balanced by optimizing the medium composition. However, here we ask the inverse question: What is the (optimal) AA composition of a hypothetical recombinant protein so that it is maximally produced under an observed growth characteristic in a known medium?

| Organsim | Model | Metabolites | Reactions | Genes | Genes/Reactions |
|----------|-------|-------------|-----------|-------|-----------------|
| *E. coli* | iJO1366 [12] | 1,136 | 2,251 | 1,366 | 0.61 |
| *S. cerevisiae* | Yeast 6 [13] | 1,458 | 1,888 | 900 | 0.48 |
| *H. sapiens* | Recon 2 [10] | 2,626 | 7,440 | 1,789 | 0.24 |
| CHO | iCHO [9] | 2,341 | 6,663 | 1,766 | 0.26 |

**Table 1**: Comparison of number of genes, metabolites and reactions for different organism-specific GSMMs.

## 2  Methods

We used the generic, non strain-specific iCHO GSMM and flux-balance analysis (FBA) [14] in combination with experimental data [15] to model a CHO-DG44 derived cell line. The cells were characterized by iCHO's (internal) stoichiometric matrix S, and subject to the usual steady-state constraint, $Sv = 0$. The *in silico* growth rate was fixed to an experimentally determined value of 0.0247 mmol/(gDW h) observed during the exponential growth phase of the fermentation [15]. Exchange reactions representing uptake or secretion of AAs, glucose, lactate, oxygen, carbon dioxide and ammonia were set to the experimental values reported by Selvarasu *et al.* [15]. Uptake rates of phosphate, protons, and water were unconstrained. Other uptake reactions, however, were all set to zero, while other secretion reactions were left unconstrained. Finally, we defined a (generic) chemical reaction for a protein of interest, POI,

$$\lambda_{ATP}ATP + \lambda_{ala}ala + ... + \lambda_{val}val \rightarrow POI \tag{1}$$
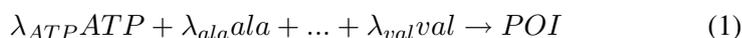
whose production was maximized in the FBA. POI required energy in form of ATP and all AAs. All components were weighted by their respective stoichiometric coefficient $\lambda_i$.

**Single AA availability.**   For each AA $i \in \{ala, ..., val\}$ we maximized Equation (1) setting $\lambda_{ATP} = 4.306$ and $\lambda_i = 1$, while the stoichiometric coefficients of all other AAs were set to zero.

**ATP limitation.**   We repeated the analysis for single AA availability (see above) with different energy requirements $\lambda_{ATP}$ ranging from 2 to 6.

**Maximum yields for typical CHO products.**   We used actual composition data for commonly produced biopharmaceuticals in CHO [1, 16, 17] and maximized Equation (1), subject to the constraints listed above.

## 3 Results and Discussion

We calculated the maximal achievable rate of production for each AA under given conditions (see Methods for details). The results are shown in Panel A and C of Figure 1. Black solid lines indicate the AA requirements for biomass formation at the experimentally measured growth rate. Grey bars indicate the variability of each AA. We observed that all AAs were available in excess with serine (*ser*) and phenylalanine (*phe*) most and least abundant, respectively. Unless the POI was extremely low on *phe*, we expected *phe* to be a metabolic bottleneck. In fact, we checked by example that for maximum production of Immunoglobulin G (IgG) of all AA, only *phe* had a non-vanishing shadow price (results not shown). Theoretically reducing the fraction of *phe* in IgG by 50% (from $\lambda_{phe} = 0.04$ to $\lambda_{phe} = 0.02$ in the protein's composition) resulted in two-fold higher production rate.

Panels B and D represent the effect of applying an energy burden onto the AA production for the same experimental conditions as previously used. For half of the AAs we observed a reduced secretion rate upon an increased ATP demand. More specifically, the production of proline (*pro*), glutamate (*glu*), glycine (*gly*), aspartate (*asp*) and alanine (*ala*) were energy limited. The production (secretion) of *ser*, asparagine (*asn*), glutamine (*gln*) and arginine (*arg*) were at least partially affected by the ATP settings. The remaining AAs were unaffected by an increasing ATP demand.

Based on the preceding analysis we expected that proteins high on *ser* and *pro* and low on *phe* will be efficiently produced. Table 2 lists the AA composition for several recombinant proteins typically produced in CHO cells. When comparing Figure 1 and Table 2, we expected a higher productivity for BMP-2, IgG and EPO than for HBsAg and hGH. This hypothesis was confirmed by optimizing for product formation in FBA with the corresponding AA composition (Table 2). More specifically, HBsAg had the highest *phe* content and therefore was produced at the lowest rate. The low *phe* content in IgG, EPO and BMP-2 resulted in higher productivity. Note, however, that high levels of *pro* and *ser* alone are not indicative of high productivity unless the *phe* content is low, as illustrated by HBsAg.

## 4 Conclusions

COBRA based approaches are commonly used to gain a quantitative insight into cellular metabolism. A prerequisite for such an analysis is the availability of a GSMM. Recent publications of sequencing data [18, 19, 20] and combination of reconstruction efforts of several research groups provided the

| POI | % phe | % pro | % ser | maximum production rate [$\mu$mol/(gDW h)] |
|---|---|---|---|---|
| EPO | 2.1 | 5.7 | 6.2 | 4.32 |
| BMP-2 | 2.6 | 6.1 | 7.0 | 3.40 |
| IgG | 3.1 | 9.9 | 9.0 | 2.85 |
| hGH | 6.8 | 4.2 | 8.9 | 1.32 |
| HBsAg | 7.1 | 10.2 | 10.6 | 1.26 |

**Table 2**: Maximum production rates for different recombinant POI as predicted by FBA. EPO, Erythropoietin; BMP-2, Bone morphogenetic protein-2; IgG, Immunoglobulin G; hGH, human growth hormone; HBsAg, hepatits B surface antigen.

basis for the generation of iCHO, currently the most up to date GSMM of CHO. Here we used iCHO to study the influence of the protein's composition on the maximum production rate of a POI under fixed nutritional conditions in a given cell line. Although the question is theoretical in nature, it may have biotechnological implications. We hypothesized that this approach could identify potential high yield producer cell lines if the cell line specific optimal AA composition matches the AA composition of the desired recombinant POI as closely as possible. In fact, our analysis revealed that under the studied conditions, proteins high on *ser* and low on phe are particularly well produced. In CHO, phe is an essential AA. It cannot be synthesized and it is therefore limited by its uptake rate. This is not the case for either *ser* or *pro*; both biosynthesis pathways are available in CHO and therefore they are flexible in covering additional demands. Thus, by matching the composition of the POI to the excess of AAs present in the cell we predicted preferential POIs.

We used FBA to select the best POI for efficient production in a given cell line under defined conditions. We first identified the limiting component and then screened for the productivity of all AAs. At the moment our approach is very simplistic in that it considers only one AA at a time. However, we are developing an optimization framework that simultaneously accounts for all AAs.

The work rests on the assumption that the growth characteristics remain unaffected by the product. In general this is known to be oversimplifying. However, for proteins similar in compositions, like BMP-2 and EPO, the assumed lack of feedback on growth might be justifiable.
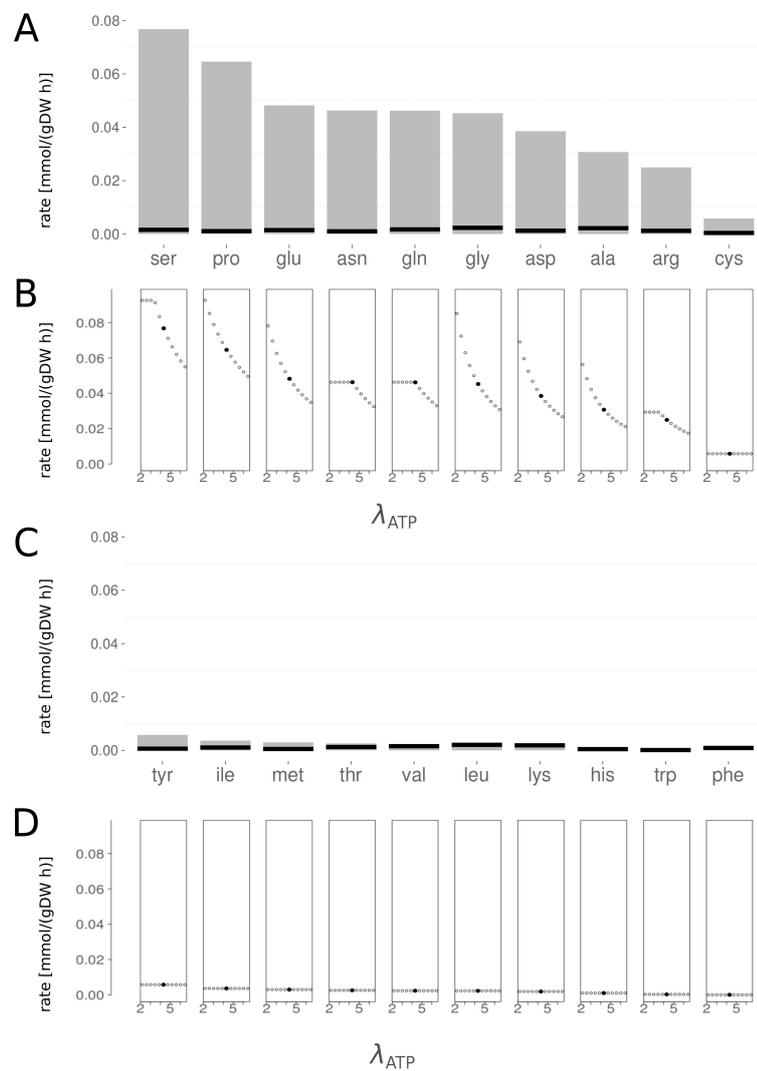
**Figure 1**: Panel A and C: Maximal production rate for each amino acid (grey bars). Black horizontal lines indicate the amount of each amino acid in the biomass of CHO using the experimentally determined growth rate in the exponential phase. Panel B and D: variation of the production rate of each amino acid (y-axis) with respect to the number of ATP moles (x-axis) needed for the reaction. The solid circles in panels B and D indicate the ATP values used for the computations in panels A and C ($\lambda_{ATP} = 4.306$). Abbreviations: $cys$, cysteine; $his$, histidine; $ile$, isoleucine; $leu$, leucine; $lys$, lysine; $met$, methionine; $thr$, threonine; $trp$, tryptophan; $tyr$, tyrosine; $val$, valine.

### References

[1] Jayapal et al. (2007). Recombinant protein therapeutics from CHO cells-20 years and counting. *Chemical Engineering Progress* **103**:10, 40.

[2] Walsh, Gary (2005). Biopharmaceuticals: recent approvals and likely directions. *Trends in Biotechnology* **23**:11, 553–558.

[3] Bort, Juan A Hernández, Beate Stern, and Nicole Borth (2010). CHO-K1 host cells adapted to growth in glutamine-free medium by FACS-assisted evolution. *Biotechnology Journal* **5**:10, 1090–1097.

[4] Huang, Yao-Ming et al. (2010). Maximizing productivity of CHO cell-based fed-batch culture using chemically defined media conditions and typical manufacturing equipment. *Biotechnology Progress* **26**:5, 1400–1410.

[5] Kim, Sung Hyun and Gyun Min Lee (2007). Functional expression of human pyruvate carboxylase for reduced lactic acid formation of Chinese hamster ovary cells (DG44). *Applied Microbiology and Biotechnology* **76**:3, 659–665.

[6] Mastrangelo, Alison J et al. (2000). Part II. Overexpression of bcl-2 family members enhances survival of mammalian cells in response to various culture insults. *Biotechnology and Bioengineering* **67**:5, 555–564.

[7] Schuetz, Robert et al. (2012). Multidimensional optimality of microbial metabolism. *Science* **336**:6081, 601–604.

[8] Winden, Wouter A et al. (2005). Metabolic-flux analysis of Saccharomyces cerevisiae CEN. PK113-7D based on mass isotopomer measurements of 13C-labeled primary metabolites. *FEMS Yeast Research* **5**:6-7, 559–568.

[9] Hefzi* , Hooman, Kok Siong Ang* , Michael Hanscho* , et al. (2016). A community built, genome-scale reconstruction of CHO cell metabolism recapitulates cell physiology and responses to bioprocess treatments. Cell Systems. Under review. * These authors contributed equally.

[10] Thiele, Ines et al. (2013). A community-driven global reconstruction of human metabolism. *Nature Biotechnology* **31**:5, 419–425.

[11] Quek, Lake-Ee et al. (2014). Reducing Recon 2 for steady-state flux analysis of HEK cell culture. *Journal of Biotechnology* **184**: 172–178.

[12] Orth, Jeffrey D et al. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism – 2011. *Molecular Systems Biology* **7**:1, 535.

[13] Heavner, Benjamin D et al. (2013). Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database* **2013**: bat059.

[14] Orth, Jeffrey D, Ines Thiele, and Bernhard Ø Palsson (2010). What is flux balance analysis? *Nat Biotech* **28**:3, 245–248.

[15] Selvarasu, Suresh et al. (2012). Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and Bioengineering* **109**:6, 1415–1429.

[16] Van Dyk, Derek D et al. (2003). Identification of cellular changes associated with increased production of human growth hormone in a recombinant Chinese hamster ovary cell line. *Proteomics* **3**:2, 147–156.

[17] Pendse, Girish J, Subhash Karkare, and James E Bailey (1992). Effect of cloned gene dosage on cell growth and hepatitis B surface antigen synthesis and secretion in recombinant CHO cells. *Biotechnology and Bioengineering* **40**:1, 119–129.

[18] Xu, Xun et al. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nature Biotechnology* **29**:8, 735–741.

[19] Brinkrolf, Karina et al. (2013). Chinese hamster genome sequenced from sorted chromosomes. *Nature Biotechnology* **31**:8, 694–695.

[20] Lewis, Nathan E. et al. (2013). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nature Biotechnology* **31**:8, 759–765.

# LIST OF ATTENDEES
(February 15$^{th}$, 2016)

AMAR Patrick                          (patrick.amar@sys2diag.cnrs.fr)
ANDRIGHETTI Tahila                    (tahilaandrighetti@gmail.com)
BALLET Pascal                         (pascal.ballet@univ-brest.fr)
BEURTON-AIMAR Marie                   (beurton@labri.fr)
BIANE Celia                           (cbiane@ibisc.univ-evry.fr)
BOUFFARD Marc                         (marc.bouffard@lri.fr)
BOUYIOUKOS Costas                     (costas.bouyioukos@issb.genopole.fr)
BUCCHINI François                     (bucchini@issb.genopole.fr)
CARBONELL Pablo                       (pablo.carbonell@manchester.ac.uk)
CARREIRA Luis                         (luis.carreira@mpi-marburg.mpg.de)
CHAVES Madalena                       (madalena.chaves@inria.fr)
CECCHI Dario                          (dario.cecchi@unitn.it)
CORNILLON Emilien                     (ecornillon@i3s.unice.fr)
COSTA Pedro                           (costapr@ibb.unesp.br)
CSIKASZ-NAGY Attila                   (attila.csikasz-nagy@fmach.it)
DI VOZZO romain                       (romaindivozzo@hotmail.com)
ELATI Mohamed                         (mohamed.elati@issb.genopole.fr)
FRANÇOIS Paul                         (paulf@physics.mcgill.ca)
GALLEGUILLOS Sarah Noel               (sarah.galleguillos@boku.ac.at)
GELENBE Erol                          (e.gelenbe@imperial.ac.uk)
GIMÉNEZ ANDRÉS Manuel                 (manugimand@gmail.com)
GLASS John                            (jglass@jcvi.org)
HERNANSAIZ-BALLESTEROS Rosa           (rosa.hernansaiz@kcl.ac.uk)
HSIN Kin-Yi                           (kunyi.hsin@oist.jp)
ISSA Razanne                          (razanne.issa@ibgc.cnrs.fr)
KARR Jonathan                         (karr@mssm.edu)
KÉPÈS François                        (francois.kepes@issb.genopole.fr)

KHAN Shawez                      (shawez.jmi@gmail.com)
KORCSMAROS Tamas                 (tamas.korcsmaros@tgac.ac.uk)
LE GALL Pascale                  (pascale.legall@issb.genopole.fr)
LEPAGE Thibaut                   (thibaut.lepage@ujf-grenoble.fr)
LÓPEZ ZAZUETA Claudia            (claudia.lopez-zazueta@inria.fr)
MADEC Morgan                     (morgan.madec@unistra.fr)
MAZAT Jean-Pierre                (jean-pierre.mazat@phys-mito.u-bordeaux2.fr)
MIRAGLIO Benjamin                (miraglio@i3s.unice.fr)
MOLINA Franck                    (franck.molina@sys2diag.cnrs.fr)
MORENO Mauro                     (mauromobel@gmail.com)
NORRIS Victor                    (victor.norris@univ-rouen.fr)
OLIVEIRA DA COSTA Noémie         (noemie.oliveiradacosta@gmail.com)
PANDI Amir                       (amrpndsmsnb@gmail.com)
PERES Sabine                     (sabine.peres@lri.fr)
PETRIZZELLI Marianyela           (Marianyela.Petrizzelli@moulon.inra.fr)
RONDELEZ Yannick                 (rondelez@iis.u-tokyo.ac.jp)
ROSATI Elise                     (elise.rosati@gmail.com)
RUBANOVA Natalia                 (nrubanova@gmail.com)
SANTOS Silvia                    (silvia.santos@imperial.ac.uk)
SCHILLER Stefan                  (stefan.schiller@frias.uni-freiburg.de)
SCHMIDT Markus                   (schmidt@biofaction.com)
SIMMEL Friedrich                 (simmel@ph.tum.de)
THIRIET-RUPERT Stanislas         (Stanislas.Thiriet.Rupert@ifremer.fr)
TRÉBULLE Pauline                 (pauline.trebulle@gmail.com)
TROSSET Jean-Yves                (jean-yves.trosset@supbiotech.fr)
ZAHARIA Alexandra                (alexandra.zaharia@u-psud.fr)
ZAWORSKI Julie                   (julie.zaworski@issb.genopole.fr)
ZELISZEWSKI Dominique            (dominique.zeliszewski@issb.genopole.fr)